THE EFFECTS OF EMOTIONS AND THEIR REGULATION ON DECISION-MAKING PERFORMANCE IN AFFECTIVE SERIOUS GAMES

Petar Jerčić

Blekinge Institute of Technology
Doctoral Dissertation Series No. 2019:06

Department of Creative Technologies

The Effects of Emotions and Their Regulation on Decision-Making Performance in Affective Serious Games

Petar Jerčić

Blekinge Institute of Technology Doctoral Dissertation Series
No 2019:06

The Effects of Emotions and Their Regulation on Decision-Making Performance in Affective Serious Games

Petar Jerčić

Doctoral Dissertation in
Game Development

Department of Creative Technologies
Blekinge Institute of Technology
SWEDEN

2019 Petar Jerčić
Department of Creative Technologies
Publisher: Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden
Printed by Exakta Group, Sweden, 2018
ISBN: 978-91-7295-370-3
ISSN:1653-2090
urn:nbn:se:bth-17557

[...] καὶ τῶν ἐπιστημῶν δὲ τὴν αὑτῆς ἕνεκεν καὶ τοῦ εἰδέναι χάριν αἱρετὴν οὖσαν μᾶλλον εἶναι σοφίαν ἢ τὴν τῶν ἀποβαινόντων ἕνεκεν [...]

[...] Again among the sciences we consider that that science which is desirable in itself and for the sake of knowledge is more nearly Wisdom than that which is desirable for its results [...]

Aristotle, *Metaphysics*, London: Heinemann, 1933, 928a 15 (translated by Hugh Tredennick and G. Cyril Armstrong)

Acknowledgements

Firstly, my sincere gratitude goes to my advisor Assoc. Prof. Veronica Sundstedt for the continuous support through my Ph.D. study and the related research, for her patience, motivation, and for providing a glimmer of hope. Her guidance helped me during the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

My sincere thanks also go to Prof. Craig Lindley and Dr. Stefan Johansson, who provided me with an opportunity to join their team and research facilities. Without their precious support and guidance, it would not have been possible to conduct this research. In particular, I am grateful to Dr. Hans Tap for his insightful comments and advice. Moreover, I wish to thank Dr. Charlotte Sennersten for her advice, as well as the possibility to develop the experiments and perform research together.

I thank my fellow colleges Johan Hagelbäck, Olle Hilborn, Wei Wen, and Henrik Cederholm, with whom I have developed the serious game artifacts and the experiments in the studies we have worked together on. My thanks also extend to our stimulating discussions, for the sleepless nights, and all the fun along the way. Also, I thank my colleagues in the other institutions: Dr. Gilbert Peffer for working on the Design and Evaluation Framework together; Dr. Philipp Astor and Dr. Marc Adam for the opportunity to work on the Auction Game and apply it to the financial research context, in the experiments we collaborated on together.

With a special mention to Johan Svensson, Malte Kirst, Dr. Kristina Schaaff, Dr. Jeanette Eriksson. It was fantastic to have the opportunity to work together with you on this research. I would also like to thank Eva Norling and her staff for their help regarding the library services.

Last but not least, I would like to thank my family and my friends for supporting me spiritually throughout this journey.

Thank you for all your encouragement!

List of Publications

Paper I

Jerčić P., Cederhold H., "The Future of Brain–Computer Interface for Games and Interaction Design", position paper for *Biosplay workshop at Fun and Games Conference, 2010.*

Paper II

Peffer G., Cederholm H., Clough G., Jerčić P., "Evaluating Games Designed to Improve Financial Capability", position paper for *ECEL 2010, 9th European Conference on e-Learning, 2010.*

Paper III

Jerčić P., Astor P. J., Adam M. T. P., Hilborn O., Schaaff K., Lindley C., Sennersten C., Eriksson J., "A Serious Game Using Physiological Interfaces for Emotion–Regulation Training in the Context of Financial Decision–Making", *Proceedings of the 20th European Conference on Information Systems (ECIS 2012).* AIS Electronic Library (AISeL), 2012.

Paper IV

Astor P. J., Adam M. T. P., Jerčić P., Schaaff K., Weinhardt, C., "Integrating Biosignals into Information Systems: A NeuroIS Tool for Improving Emotion–Regulation", *Journal of Management Information Systems*, 2013, 30(3), pp. 247-278.

Paper V

Jerčić P., Sennersten C., Lindley C., “Modeling Cognitive Load and Physiological Arousal Through Pupil Diameter and Heart Rate”, In: *Multimedia Tools and Applications journal*, Springer, 2018.

This journal publication is the extended conference publication from Jerčić P., Sennersten C., Lindley C., “The Effect of Cognitive Load on Physiological Arousal in a Decision–Making Serious Game”, In: *2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, Athens, 2017, pp. 153-156.

Paper VI

Jerčić P., Wen W., Hagelbäck J., Sundstedt V., “The Effect of Emotions and Social Behavior on Performance in a Collaborative Serious Game Between Humans and Autonomous Robots”, In: *International Journal of Social Robotics*, 2018, 10(1), pp. 115–129.

Paper VII

Jerčić P., Hagelbäck J., Lindley C., “Physiological Affect and Performance in a Collaborative Serious Game Between Humans and an Autonomous Robot”, submitted to a journal.

This journal publication is the extended conference publication from Jerčić P., Hagelbäck J., Lindley C., “Physiological Affect and Performance in a Collaborative Serious Game Between Humans and an Autonomous Robot”, In: *Entertainment Computing – ICEC 2018. Lecture Notes in Computer Science*, 2018, vol. 11112, pp. 127–138.

Paper VIII

Jerčić P., Sundstedt V., “Practicing Emotion–Regulation Through Biofeedback on the Decision–Making Performance in the Context of Serious Games: A Systematic Review”, , In: *Entertainment Computing* journal, 2019, vol. 29, pp. 75–86.

Abstract

Emotions are thought to be one of the key factors that critically influence human decision–making. Emotion–regulation can help to mitigate emotion–related decision biases and eventually lead to a better decision performance. Serious games emerged as a new angle introducing technological methods to practicing emotion–regulation, where meaningful biofeedback information communicates player's affective states to a series of informed gameplay choices. These findings motivate the notion that in the decision context of serious games, one would benefit from awareness and regulation of such emerging emotions.

This thesis explores the design and evaluation methods for creating serious games where emotion–regulation can be practiced using physiological biofeedback measures. Furthermore, it investigates emotions and the effect of emotion–regulation on decision performance in serious games. Using the psychophysiological methods in the design of such games, emotions and their underlying neural mechanism have been explored.

The results showed the benefits of practicing emotion–regulation in serious games, where decision–making performance was increased for the individuals who down–regulated high levels of arousal while having an experience of positive valence. Moreover, it increased also for the individuals who received the necessary biofeedback information. The results also suggested that emotion–regulation strategies (i.e., cognitive reappraisal) are highly dependent on the serious game context. Therefore, the reappraisal strategy was shown to benefit the decision–making tasks investigated in this thesis. The results further suggested that using psychophysiological methods in emotionally arousing serious games, the interplay between sympathetic and parasympathetic pathways could be mapped through the underlying emotions which activate those two pathways. Following this conjecture, the results identified the optimal arousal level for increased performance of an individual on a decision–making task, by carefully balancing the activation of those two pathways. The investigations also validated these findings

in the collaborative serious game context, where the robot collaborators were found to elicit diverse affect in their human partners, influencing performance on a decision–making task. Furthermore, the evidence suggested that arousal is equally or more important than valence for the decision–making performance, but once optimal arousal has been reached, a further increase in performance may be achieved by regulating valence. Furthermore, the results showed that serious games designed in this thesis elicited high physiological arousal and positive valence. This makes them suitable as research platforms for the investigation of how these emotions influence the activation of sympathetic and parasympathetic pathways and influence performance on a decision–making task.

Taking these findings into consideration, the serious games designed in this thesis allowed for the training of cognitive reappraisal emotion–regulation strategy on the decision–making tasks. This thesis suggests that using evaluated design and development methods, it is possible to design and develop serious games that provide a helpful environment where individuals could practice emotion–regulation through raising awareness of emotions, and subsequently improve their decision–making performance.

Contents

1 Introduction **1**

2 Background **7**

2.1 Models of Emotions . . . 7
2.2 Serious Games . . . 9
2.3 Psychophysiology . . . 11
2.4 Affective Computing . . . 12
2.5 Brain–Computer Interface . . . 14
2.6 Biofeedback . . . 16
2.7 Emotion–Regulation . . . 19
2.8 Decision–Making . . . 22
2.9 Human–Robot Interaction . . . 24
2.10 Technological Implementation . . . 26
2.10.1 Electroencephalography . . . 26
2.10.2 Galvanic Skin Response . . . 27
2.10.3 Electrocardiography . . . 29
2.10.4 Pupil . . . 32
2.10.5 Electromyography . . . 34
2.11 Related Work . . . 34
2.11.1 Serious Games in the Field of Healthcare and Well–Being . 34
2.11.2 Serious Games using Biofeedback for Emotion–Regulation Training . . . 35
2.11.3 Serious Games in the Context of Financial Decision–Making 35
2.11.4 Serious Games in the Field of Human–Robot Interaction . . 36
2.11.5 Serious Games for Practicing Emotion–Regulation through Biofeedback on the Decision–Making Performance . . . 37
2.11.6 Summary . . . 38

3 Approach **39**

3.1 Aim and Scope 39
3.2 Research Questions 40
3.3 Methodology 42
3.3.1 The Implementation of Serious Games 42
3.3.2 Experimental Setup 46
3.3.3 Experiment Procedure 50
3.3.4 Data Collection 51
3.3.5 Data Reduction and Analysis 51
3.4 Threats 54
3.5 Ethics 55
3.6 Contributions 56
3.6.1 Summary of Contributions 62
3.7 Discussion 65
3.7.1 Development of the Serious Games 68
3.8 Limitations 70
3.9 Conclusion 71
3.10 Future Work 74

4 The Future of Brain-Computer Interface for Games and Interaction Design 77
4.1 Introduction 78
4.2 Epoc for Game Design 78
4.3 Epoc for Interaction Design 79

5 Evaluating Games Designed to Improve Financial Capability 81
5.1 Introduction 82
5.2 The Design and Evaluation Framework 82
5.3 A Design and Evaluation Toolkit for Financial Capability Games . 83
5.3.1 Financial Capability DE tool 84
5.3.2 Behaviour Change DE tool 84
5.3.3 Game Design DE tool 85
5.3.4 Learning with Technology DE tool 86
5.4 Conclusions 87

6 A Serious Game Using Physiological Interfaces for Emotion Regulation Training in the Context of Financial Decision-Making 89
6.1 Introduction 91
6.2 Emotions, Emotion–Regulation and Decision–Making 93
6.3 Psychophysiological Measurement of Emotions 96
6.4 Game Description 97
6.4.1 Underlying Principle 97
6.4.2 Game Concept and Gameplay 97

6.4.3 Game Mechanics 101
6.4.4 Game Difficulty 101
6.5 Evaluation 103
6.5.1 Functionality and Heuristic Evaluation 103
6.5.2 Playtesting Evaluation Method 103
6.5.3 Playtesting Evaluation Results 104
6.6 Discussion and Conclusion 105

7 Integrating Biosignals into Information Systems: A NeuroIs Tool for Improving Emotion Regulation 107
7.1 Introduction 108
7.2 Conceptual Foundations 109
7.2.1 Emotion and Cognition in Financial decision–making 109
7.2.2 Emotion–Regulation 111
7.2.3 Improved Financial decision–making through Enhanced Emotion–Regulation 112
7.2.4 NeuroIs and Design Science Research 113
7.3 Design 115
7.3.1 Requirements 115
7.3.2 Concept of the Learning Environment 116
7.3.3 Rewarding Emotion–Regulation 118
7.3.4 Arousing Game Elements 118
7.3.5 Biofeedback 120
7.4 Implementation 121
7.4.1 Game Architecture 121
7.4.2 Computation of Arousal Values 122
7.4.3 Display of Arousal Values 123
7.5 Evaluation 124
7.5.1 Experimental Design 124
7.5.2 Measures 125
7.5.3 Experimental Procedure 126
7.6 Evaluation Results 126
7.6.1 Evaluation Results for Requirement R1 127
7.6.2 Evaluation Results for Requirement R2 128
7.6.3 Evaluation Results for Requirement R3 130
7.7 Discussion and Conclusions 132
7.7.1 Summary of Results 132
7.7.2 Managerial Implications 132
7.7.3 Limitations and Future Research 134
7.8 Conclusions 135
7.9 Acknowledgments 135

7.10 Appendix 136

8 Modeling Cognitive Load and Physiological Arousal Through Pupil Diameter and Heart Rate 139
8.1 Introduction 140
8.2 Background 142
8.2.1 Eye Pupil and Responses 143
8.2.2 Heart and Physiological Arousal 144
8.3 Methodology 145
8.3.1 Participants 145
8.3.2 Experimental Setup 145
8.3.3 Experimental Task 148
8.3.4 Protocol 149
8.3.5 Data Analysis 150
8.4 Results 152
8.5 Discussion 152
8.6 Limitations 153
8.7 Conclusions 154

9 The Effect of Emotions and Social Behavior on Performance in a Collaborative Serious Game Between Humans and Autonomous Robots 157
9.1 Introduction 159
9.1.1 Related Work 161
9.1.2 Human-Robot Interaction 161
9.1.3 Serious games 162
9.1.4 Emotions 163
9.1.5 Social Interaction 163
9.2 Research Questions 165
9.3 Methodology 166
9.3.1 Participants 166
9.3.2 Compliance with Ethical Standards 166
9.3.3 Experimental Setup 166
9.3.4 Hardware System 171
9.3.5 Software Platform 171
9.3.6 Experiment Procedure 172
9.3.7 Data Processing 173
9.4 Results 173
9.4.1 Collaborator conditions and performances 174
9.4.2 Elicited emotions and performances 175
9.4.3 Social presence and performances 176

9.5 Discussion 178
9.5.1 Limitations 179
9.6 Conclusion 180
9.7 Appendix A - Interactive Experience Questionnaire 182
9.8 Appendix B - Trust and Respect Questionnaire 183

10 Physiological Affect and Performance in a Collaborative Serious Game between Humans and an Autonomous Robot 187
10.1 Introduction 188
10.2 Related Work 190
10.2.1 Affect Detection 191
10.2.2 Serious Games 191
10.3 Hypothesis 192
10.4 Methodology 194
10.4.1 Participants 194
10.4.2 Experimental Setup 195
10.4.3 Experiment Procedure 197
10.4.4 Data Collection 199
10.4.5 Data Reduction and Analysis 199
10.4.6 Hardware and Software Systems 201
10.5 Results 203
10.5.1 Collaborator Conditions and Performance 204
10.5.2 Physiological Arousal 205
10.5.3 Physiological Valence 206
10.5.4 Limitation 207
10.6 Discussion 208
10.7 Conclusion 209

11 Practicing Emotion–Regulation Through Biofeedback on the Decision–Making Performance in the Context of Serious Games: A Systematic Review 213
11.1 Introduction 215
11.2 Related Work 216
11.2.1 Decision–Making 217
11.2.2 Emotion–Regulation 218
11.2.3 Biofeedback 218
11.2.4 Physiology 220
11.2.5 Physiological and Affective Computing 221
11.2.6 Serious Games 222
11.2.7 Skill–Transfer 223
11.3 Method 224

11.4 Results 229
11.4.1 Biofeedback 229
11.4.2 Tasks 232
11.4.3 Emotion–Regulation 234
11.4.4 Valence 235
11.4.5 Healthcare 236
11.4.6 Methodology 236
11.5 Discussion 237
11.6 Limitations 239
11.7 Future Work 239
11.8 Conclusion 240

References **243**

List of Acronyms

ANS	Autonomic Nervous System
BCI	Brain–Computer Interface
CNS	Central Nervous System
DE	The Design and Evaluation Framework
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
fEMG	Facial Electromyography
GBL	Game–Based Learning
GEW	Geneva Emotion Wheel
GSR	Galvanic Skin Response
HCI	Human–Computer Interaction
HHI	Human–Human Interaction
HR	Heart Rate
HRI	Human–Robot Interaction
HRV	Heart Rate Variability
IAPS	International Affective Picture System
LF/HF	Ratio of Low and High Frequency Powers
NS-SCR	Non–Specific Skin Conductance Responses
PNS	Parasympathetic Nervous System
PD	Pupil Diameter
RSA	Respiratory Sinus–Arrhythmia
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SDNN	Standard Deviation of Normal–to–Normal Heartbeat Intervals
SNS	Sympathetic Nervous System
ToH	The Tower of Hanoi

Chapter 1

Introduction

The concept of *emotion* still presents a great debate amongst scientists. James (1884) presented the still unanswered question: What is an emotion? After all this time the answer has not emerged or converged, with little agreement on where emotion stops and its causes and consequences begin. According to Lang (1995), emotions are dispositions, or states of readiness, that help people and organisms to interact with the environment. Such emotions facilitate our interactions with the socioeconomic environment; they foster meaningful interpersonal interactions, prepare behavioral responses, and enable us to take advantageous decisions (Gross, 2008). There is no clear answer to what is and what is not an emotion. Cole et al. (2004, p. 330) captured the nature of emotions in a statement: "Emotions are powerful, elusive, dynamic processes. They have the capacity to regulate other processes and to be regulated". Unfortunately, however, emotions are not always accurately processed. Goldstein (1996) discusses how emotions can get 'out of control', wherein the heat of the moment, human decision makers can be overwhelmed by their emotions and "under the influence of high levels of arousal" lose control over their actions.

Emotions do not always impair *decision–making*. They may also have a positive influence on decisions and facilitate them. The experience of emotions may have a positive impact on performance, as it is a necessary prerequisite for optimal decision–making (Bechara and Damasio, 2005). Emotions can contain relevant information, as concluded by Bechara and Damasio (2005) in their somatic marker hypothesis. This hypothesis states that emotional responses to information events (somatic markers) guide one's focus of attention. The somatic marker theory hence states that decisions are aided by emotions in the form of bodily states (Naqvi et al., 2006). It is vital for decision makers and for the organizations they represent, to adequately process emotions because they mainly affect important decisions. However, there are strong interpersonal differences in the capabilities of adequate emotional processing (Lo et al., 2005; Lo and Repin, 2002). It has been shown that those decision makers with high *emotion–regulation* capabilities perform better in trading (Fenton-O'Creevy et al., 2012b). In contrast, decision makers with low emotion–regulation capabilities frequently are more prone to taking disadvantageous decisions with undesirable consequences. Therefore, emotion–regulation can help to mitigate emotion–related decision biases and eventually lead to better decision performance. However, how can decision makers practice to improve their emotion–regulation capabilities? Unfortunately, traditional approaches seem to fail when it comes to improving emotion–regulation capabilities (Fenton-O'Creevy et al., 2012a).

Emotion–regulation can only be actively practiced when users actually face an arousing environment (Dror, 2008; Burgoon and Bonito, 2002). Emotion–regulation using *biofeedback* has been extensively investigated. However, it is only by the introduction of biofeedback where emotional states are presented on–line coupled to *serious games* with immediate feedback regarding decision–making tasks, that the role of emotion–regulation in making decisions could be put in the context (Yannakakis et al., 2016). This interactive approach to such investigation is highly context–dependent, where emotion–regulation skill–acquisition and transfer are 'maximized' when individuals are provided with engaging context–dependent serious games, which have emerged as a perfect tool that fits this purpose (Sliwinski et al., 2017).

Recent years have witnessed an increase in biofeedback integration in serious games for a number of purposes, ranging from gameplay interaction mechanisms to healthcare benefits (Yannakakis et al., 2016). Self–regulation in the face of today's high–paced stressful environment helps to cope with demanding situations, reduce negative health outcomes, and increase the overall quality of life (Parnandi and Gutierrez-Osuna, 2017). Biofeedback allows users to visualize specific physiological parameters in order to raise awareness and improve emotion–regulation so that the aversive behavior may be changed (Repetto et al., 2013). However, these visualizations are non–intuitive to many users. Biofeedback and a number of other methods to teach self–regulation, including cognitive behavior therapy, yoga, and meditation, have been validated in the clinical field as methods for tackling psychological issues that are hard to perceive by individuals (Sliwinski et al., 2017). On another hand, the individuals engaged in these methods would need to invest a significant amount of time in training or need the supervision of a therapist, which can be time and cost prohibitive. Compared to traditional approaches, serious games are interactive media which can provide a context for the on–line perception of emotional responses through biofeedback, which reflects the changes in the individuals' physiology (Yannakakis et al., 2016). Serious games emerged as a remedy to traditional approaches, providing a new angle introducing technological methods to acquiring and practicing skills. Serious games are games used for purposes other than mere entertainment (Susi and Johannesson, 2007). Corti (2006) points out an obvious advantage of serious games in allowing users to experience situations that are impossible in the real world for reasons of safety, cost, time or logistics. Serious games can have positive impact on the development of a number of different skills (Ellis et al., 2006; Corti, 2006; Van Eck, 2006; Mitchell and Savill-Smith, 2004; Squire and Jenkins, 2003; Rieber, 1996). Having those different skills as defined learning outcomes, one can see why serious games are considered as *game–based learning* applications (Corti, 2006). Such biofeedback methods provide a shortcut to raising awareness of internal bodily states and emotions, which are one of the key factors that critically influence human decision–making (Adam and Gamer, 2011; Loewenstein, 2000). Emotions and decision–making biases are even beyond conscious awareness, and it is therefore highly interesting to regulate such emotional dispositions (Fenton-O'Creevy et al., 2011). Therefore, an awareness of emotions is a necessary prerequisite for deciding to follow or not an emotional response (Gross, 1998b).

Games are a series of interesting choices where decision–making could be introduced through the design of those choices (Sütterlin et al., 2011). Furthermore, they might provide a platform where psychophysiology can be implemented as methods of biofeedback to support practicing and rewarding emotion–regulation (Parnandi and Gutierrez-Osuna, 2017; Lang and Twentyman, 1976). Such design could benefit serious games as those choices could be informed about emotional states on–line though individuals' physiology. The benefits of practicing emotion–regulation are largely dose–dependent, and this view is missing in the previous research investigations (Culbert, 2017). The author states that the individuals who regularly practice benefit both in the short– and long–term. The authors further state that serious games could mitigate these issues, but they need to take into account the player–centered design of such games. Furthermore, they can parallel the qualities of the traditional approaches regarding beneficial effects on attention and emotional regulation, while providing valuable conditions to improve accessibility, for example by giving clear instructions and multisensory feedback (Sliwinski et al., 2017).

The game design is defined as designing experiences for players, and it is an integral part of the game development process (Squire, 2006). An early approach to designing games promoted a 'one color suits all' premise, where a design solution was trying to captivate a wide player audience. To date, this methodology has been pervasive in games (Cornett, 2004). In recent years there was a paradigm shift towards player–centered game development to increase the perceived value (Chiou and Wong, 2008; Charles et al., 2005). Such game development takes into consideration individual player information to provide an enhanced gameplay experience (Charles and Black, 2004). Further evolution in the field introduced this technology to other areas, such as education and awareness of emotions in individual players.

This thesis explores game and sensor technologies that could be used to support emotion–regulation training and therefore improve decision–making. This approach investigates the role and underlying mechanism of emotions in the decision–making process. These research insights supported the development of two serious game approaches for training and investigation of decision–making. A need to create those approaches in an iterative design and development fashion introduces to the *Design and Evaluation* (De) framework which proposes practices and methods for the evaluation of serious games. The interdisciplinary process of creating such technology–enhanced serious game tools requires such a flexible framework supporting the complexities of serious game development and evaluation process, linking goals acquired from empirical studies and evaluating the tool's potential to meet those goals (Jonassen et al., 2003).

A possible way to further improve an individual's motivation on the task is to develop serious games that involve other physical players to collaborate with, which are adaptable to individuals with diverse abilities (Hocine and Gouaich, 2013). Physical robot collaborators can have direct access to the game–state and physiological information of the human partners. Furthermore, previous investigations have shown that the engaging physical non–humanoid robot collaborators can elicit emotional responses (Fiore et al., 2013; Scholl and Tremoulet, 2000), which in turn might influence human decision–making and performance on a game task (Adam and Gamer, 2011; Chanel, 2009; Shiv et al., 2005). Researchers working with *human–robot interaction* (Hri) are investigating robots as peers and colleagues with a variety of social and emotional abilities (Norman, 2003; Breazeal, 2002). In Hri–enhanced serious games, one is playing together with a physical entity eliciting diverse behavior and stronger emotional response in players, which might support higher motivation and better skill–acquisition (Hocine and Gouaich, 2013). Understanding emotional and social cues underlying such an interaction from the human perspective, it would be possible to design better serious games using helpful and intelligent robots which act according to behavioral patterns that humans can understand and relate to (Xin and Sharlin, 2006).

Current research is exclusively concerned with an individual's affective experience and motivation in serious games (Carter et al., 2014). On another hand, the authors argue that the field of serious games is already recognized in research where decision–making questions are a burning issue. Therefore, this thesis investigates the effects of practicing emotion–regulation through biofeedback on decision–making performance in the context of serious games. Furthermore, previous investigations argue that future research should examine which game features are most effective in promoting engagement and supporting skill–acquisition (Boyle et al., 2016). Therefore, this thesis focuses on the design investigation of promoting engagement and supporting skill–acquisition of emotion–regulation through an engaging and arousing environment (Dror, 2008; Burgoon and Bonito, 2002). Furthermore, this thesis introduces and evaluates a game development process to meet the mentioned design goals, where robot collaborators have been considered to further the mentioned aims. It could be possible to develop serious games using robot collaborators that could be motivating partners for playing and for practicing emotion–regulation. Furthermore, such robot collaborators could be an adequate stimulus to elicit emotional affect in human players (Pereira et al., 2012). Nevertheless, prior to such investigations with human participants, specific legal and ethical issues had to be considered. These include risk assessment and informing the participants with complete information on the studies' goals, experimental conditions and signing the informed consent (Walters et al., 2005; Crown Copyright, 2003). These were carried out in all of the studies presented. The methods investigated in this thesis should lead to developing better serious games, one with an engaging and arousing environment where the emotion–regulation skill could be acquired and practiced.

Chapter 2 presents the background and terminology. In Chapter 3, Section 3.1 gives the aim and scope of the thesis, while Section 3.2 lists the research questions. Section 3.3 details the methodology, where validity threats are discussed in Section 3.4. Section 3.5 gives a discussion on ethics. Section 3.6 concerns the contributions of the publications, as well as the author's involvement. The results are discussed in Section 3.7 with limitations presented in Section 3.8 and concluded in Section 3.9 respectively. Section 3.10 presents points for future work. Finally, the publications are presented in Chapters 4–11.

Chapter 2

Background

Emotions enable us to prepare behavioral responses and take advantageous decisions (Gross, 2009; Seo and Barrett, 2007). It is generally accepted that bodily states shape emotions and this mechanism is an important method of self–evaluation in the mindfulness approach (Fenton-O'Creevy et al., 2012a). The authors state that if those bodily states are measured through physiological responses, it is possible to provide objective quantitative insight into the elusive concept of emotions. Research in psychophysiology suggests that the measurement of physiological responses could quantitatively characterize a number of emotional states (Nacke and Lindley, 2008). This makes *psychophysiology* a useful research tool for evaluating: game experience, motivation, emotion–regulation, or decision performance in games. On a conscious level, one interprets those physiological reactions in regards to the external stimuli, which give rise to the subjective experience of emotions in one's mind (D'Mello et al., 2013). This kind of information can also be qualitatively probed using questionnaires. Both approaches to studying emotions have been used and compared in this thesis.

2.1 Models of Emotions

Various overlapping concepts within the models of emotions and affect are translated through psychology to computer science, especially its determinants, measures and how these relate to emotion–regulation. This makes it difficult to develop valid measures and a common understanding of emotions and affect, especially in the context of decision–making in serious games. Russell (1980) gen-

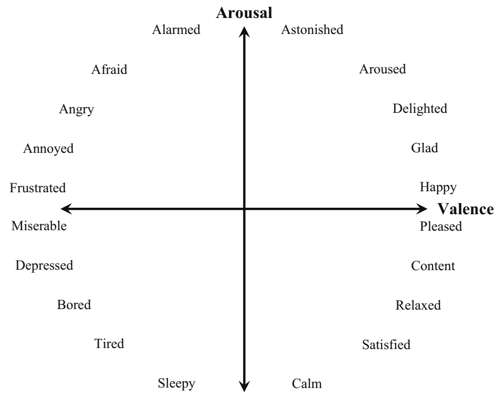

Figure 2.1: Emotions in the valence–arousal space. Adapted from Russell (1980).

erally classified emotions by their independent components, *arousal* and *valence*. According to his circumplex model, arousal represents the level of excitement whereas valence defines whether the current emotional state is positive or negative, as illustrated in Figure 2.1. Thus, emotions and their measurements contain a combination of both valence and arousal. Physiology (e.g., *heart rate* (HR), *galvanic skin response* (GSR)) provides an opportunity to extract and interpret both valence and arousal (Leng et al., 2007; Anttonen and Surakka, 2005; Cacioppo et al., 1986) in relation to emotional stimuli. Nevertheless, Winton et al. (1984) state that arousal measured through HR has been the primary interest in most of the emotion–regulation serious games studies, as argued in Chapter 11.

Emotions are indicators of the game experience in players (Nacke and Lindley, 2008). They motivate cognitive decisions made during gameplay, so it is only natural that game designers are seeking methods to understand better the positive and negative emotions that one feels when one is playing games (El-Nasr et al., 2013). There is a possibility that emotions also hinder advantageous decision–making, especially if they are not accurately processed and get 'out of control' (Adam et al., 2013; Bechara and Damasio, 2005; Shiv et al., 2005; Loewenstein, 1996). Individuals can be overwhelmed by their emotions in demanding situations and lose control over their actions, under the influence of high levels of arousal (Astor et al., 2013b). This may significantly threaten rational cognition and decision–making (Kuhnen and Knutson, 2005). Therefore, there is a need to raise awareness in individuals that emotional processes can affect their decision performance.

2.2 Serious Games

Validated biofeedback self–guided interventions (i.e., meditation, cognitive–based therapy, and yoga) suffer from high dropout rates due to the unengaging nature of exercises and lack of motivation (Rose et al., 2013; Davis and Addis, 1999). Furthermore, these techniques teach self–regulation in quiet, controlled settings, which may not generalize to the context of real–world stressors (Driskell and Johnston, 1998). Traditional approaches for practicing emotion–regulation fail mostly because of the limited capacity for self–monitoring (Astor et al., 2013b). As traditional approaches fail in improving emotion–regulation skill–acquisition, serious games emerged as a new angle introducing technological methods to practicing emotion–regulation, where meaningful biofeedback information communicates players' emotional states on–line (Astor et al., 2013b; Fenton-O'Creevy et al., 2012a). Furthermore, serious games set in the relevant context close to the real–world stressors promote higher motivation and engagement in players, which might lead to better skill–acquisition.

Serious games can be understood as games that aim at purposes other than entertainment alone since they are engaged in a 'serious' investigation of certain aspects of human endeavors (Liu et al., 2013). Serious games that are repeatable, highly engaging, and motivating may result in improved training activities; furthermore, they may have positive effects on the development of new knowledge and skills (Léger et al., 2011; Anderson and Lawton, 2009). A distinct advantage of serious games is to allow individuals to experience situations that are impossible in the real world (Corti, 2006). The author further states that serious games need to captivate and engage the individual for a specific purpose, such as to be challenging enough to elicit an emotional response. Serious games can be understood in terms of their rules, interfaces, and the concept of play that they deploy. Interaction and feedback are achieved through virtual or physical actors that interact with their players. From an affective perspective, such interaction interfaces require an on–line affective system analyzing the physiological activity of the players to recognize and map their emotions, and present biofeedback. For example, evidence shows that salivary cortisol is highly reliable for assessing arousal, but it is not suited for providing users with an on–line biofeedback (Riedl et al., 2010).

The game–development research uses game metrics data to investigate players' behavior, by logging changes in the game system (Guardini and Maninetti, 2013). Contrary to the traditional behavioral gameplay metrics where a player is able to fake doing an activity while cognitively engaging on another, physiology allows for an objective measure of arousal and valence during gameplay sessions, measured continuously and unobtrusively in players (Nacke et al., 2009). It allows for the minimally–biased assessment of how a player is reacting to gameplay events, which informs game designers if they succeeded in causing an emotional response in players with the game decisions they designed (Nacke, 2013). *Game metrics* can be very useful for logging players' behaviors, but they come short at explaining this behavior or the resulting player experience. Such metrics can identify some potential issues with a game, but for game developers to know what to change in order to improve their design, one should aim to understand why it was an issue in the first place. *Player metrics* are derived or collected from players, as opposed to players' gameplay (game metrics), which are collected from games (McAllister et al., 2013). On another hand, asking players to self–report (e.g., interviews or questionnaires) relies on their awareness, recall, and cognitive filtering abilities to function before a response emerges, and probably a biased one (McAllister et al., 2013). Authors further state that players cannot accurately remember their gameplay experience, even after short game sessions. In psychology, this is known as the *serial position effect* (Feigenbaum and Simon, 1962). The distinction between emotional states and game events that caused them is possible when analyzing psychophysiological reactions together with game metrics data, to pinpoint the exact game events and contextualize physiological reactions of players (Kivikangas et al., 2011; Nacke et al., 2008). Physiological responses of players depend on the context of games. Thus designing a game context in which players experience their emotions is as important as designing game–related cues that trigger those emotions (Nacke, 2013). They become meaningful only when analyzed using the correct game design context and correct signal processing procedures (Nacke, 2009; Mandryk, 2008). This concept is an important tool in game evaluation studies.

Using this new body of information, many serious games and applications in the domain of clinical medicine today use biofeedback to help people deal with emotional problems such as anxiety, phobia and post–traumatic stress. It is reasonable to expect that games which give instantaneous feedback on players' emotional states provide a valid pedagogical approach since direct feedback has been shown to be beneficial in task training situations (Kluger and DeNisi, 1996). Poorly developed games are unappealing, demotivating, and are unlikely to generate useful skill–acquisition in individuals. As Malone (1980) states, a training environment has to be optimally complex where individuals know enough to have expectations about what will happen, but where these expectations are sometimes unmet. Providing some performance feedback to clearly explain why expectations are not the ones achieved, motivates individuals for deeper understanding. Malone (1980) further states that intrinsic motivation through gameplay provides encouragement to spend more time and effort practicing skill–acquisition, feel better about it, and use it more in the future. Moreover, such enjoyable experiences stimulate or encourage players to intensely engage with content, more than traditional methods. Therefore, serious games often deliver skill–acquisition in simulated environments where physical environments are inaccessible, unavailable, too expensive or dangerous (Ritterfeld et al., 2009; Corti, 2006).

The neuroscience perspective might explain the skill–acquisition through using biofeedback in serious games, where performing goal–directed tasks (i.e., playing a serious game) leads to dopamines release in the striatum of the brain (Koepp et al., 1998). The authors found that there is a monotonic increase in striatal dopamine levels released during gameplay (compared to a baseline) and that the effect was sustained after the gaming session ended. Such dopamine release is an indicator of memory storage events, attention, and is also involved in learning stimuli or actions that predict rewarding or aversive outcomes (Achtman et al., 2008). Therefore, serious games may facilitate improved awareness of arousal and effective skill–acquisition/transfer of emotion–regulation (Parnandi and Gutierrez-Osuna, 2017). O'Donohue and Fisher (2009) propose that the basic principle of effective emotion–regulation skill–transfer is: a) teach and explain the regulation strategy; b) practice regulation skill in a simple situation; c) progress to more and more challenging situations. For these emotion–regulation principles to work, an environment that induces stress and arousal is required, which may be difficult to recreate in academic settings or role plays in a briefing room (Bouchard et al., 2012). The investigations using such environments have to take into consideration specific legal and ethical issues already discussed in the Introduction. Serious games for practicing emotion–regulation through biofeedback on the decision–making performance continuously display individuals' emotional states and adapts the difficulty of the task to these states, which supports skill–acquisition of down–regulating high levels of arousal (Astor et al., 2013b).

2.3 Psychophysiology

Psychophysiology is concerned with the physiological bases of psychological processes, as a perspective of studying the interface of subjective states and physiological processes. Psychophysiological measures are often used to study emotions, emotional states and attentional responses to stimuli during exertion, to better understand cognitive processes (Vyzas and Picard, 1998; Cacioppo and Tassinary, 1990; Ekman et al., 1983). The concept of *physiological arousal* has been generally validated in models of emotion (Morrison, 2001; Porges, 1995; Cannon, 1994). With examples being the pupil diameter and GSR, influenced by physiological arousal, together with the attention and interest (Andreassi, 2007; Lang et al., 2005; Haag et al., 2004; Stern et al., 2001; Plutchik, 1984). Physiology was employed to measure emotional states of humans in response to robots with more than 80% success rate (Rani et al., 2007; Kulic and Croft, 2006; Picard et al., 2001).

2.4 Affective Computing

Affective computing is defined as “computing that relates to, arises from, or deliberately influences emotions” (Picard, 2000, p. 3). In this regard, Picard et al. (2001) stressed the importance of software intelligence having the ability to sense and respond to the users’ affective states based on physiological measurements. Affective computing is a branch of science which studies psychophysiology and its application in games, autonomous robotic systems or software applications. Analysis of users’ psychophysiological biosignals creates an interpretation model of the users’ affect, and such a model could be communicated inside of a game, or any other autonomous robotic systems or software application. *Physiological computing* represents a category of affective computing that incorporates dynamic software systems adapting themselves in real–time to the psychophysiological activity of their users (Fairclough, 2009). The main goal of physiological computing is to build a computer that responds to user emotion, cognition and motivation. Moreover, physiological measures were used in previous research to investigate how interacting with information technology can induce considerable levels of stress in the users (Riedl et al., 2012; Zhai and Barreto, 2006a). Humans engaged in interaction expect recognition of affective states (Picard, 2000). Such serious games or systems could be aware of the players’ emotional states and could provide meaningful biofeedback information. Psychophysiological methods draw theoretical background from cognitive science (neuroscience, neuroengineering) and affective computing draws theoretical background from both psychophysiology and cognitive science.

Cognitive science is an interdisciplinary scientific study of the mind and its processes regarding cognition, perception, language, memory, reasoning and emotion (Friedenberg and Silverman, 2005). Furthermore, it also explores how information is represented and transformed in human/animal behavior, nervous systems or machines. As the field is highly interdisciplinary, research often cuts across multiple areas of science, drawing on research methods mentioned in the previous paragraph. By measuring behavioral responses to different stimuli (using psychophysiology), one can deepen the understanding of a variety of cognitive processes and how those stimuli are processed.

While *neuroscience* is a scientific study of the nervous system, *cognitive neuroscience* is an academic field concerned with the scientific study of biological substrates underlying cognition, with a specific focus on the neural substrates of mental processes (Squire et al., 0012). It explores the underlying neural mechanism of psychological/cognitive functions in the brain. At the cognitive level, cognitive neuroscience addresses questions of how psychological functions are produced by neural circuitry. Neural imaging and sophisticated experimental techniques from cognitive psychology allow addressing abstract questions such as how human cognition and emotion are mapped to specific neural substrates.

More recently, psychophysiologists have been interested in the *central nervous system* (CNS), exploring cortical brain potentials, brain waves, and utilizing advanced technology. By measuring physiological valence and arousal, their independence in forming emotions are assumed. However, as activation of *parasympathetic nervous system* (PNS) and *sympathetic nervous system* (SNS) is the underlying mechanism of emotions, one can only be assured of their dependence. Thus, one needs to point out that physiological valence and arousal are a model of emotions, allowing exploration of this complex and interdependent phenomena (Steinhauer et al., 2004).

Studying the relationship between brain and behavior is investigated through physiological signals, which are guided by activity in the *autonomic nervous system* (ANS) (Hugdahl, 1995). The author further states that this investigation informs researchers on what a mind is doing at the moment of a particular game event. This makes psychophysiology a useful tool for evaluating game development and player experience (i.e., excitement, emotion or mental workload) in games through conducting experiments (Nacke, 2013). A significant part of the player experience arises from emotional reactions (Järvinen, 2009). Neuroscience demonstrates that on the internal level of automatic physical reactions emotions are created by the same neural circuitry of ANS, interpreted as different by the cognitive appraisal of the affective experience based on the context cues (Posner et al., 2005; Watson et al., 1999; Lazarus, 1991).

Cognitive psychology and psychophysiology are closely related to the field of neuroscience, which primarily concerns itself with relationships between psychological events and brain responses. The use of theories, methods, and tools offered by cognitive neuroscience and psychophysiology in the science of designing information systems is still in its infancy (Ortiz De Guinea et al., 2013). vom Brocke et al. (2012) propose three strategies for the framework on how to use those tools and theories: a) inform the building and evaluation of information system artifacts; b) use of neuroscience tools to evaluate information system artifacts; c) use neuroscience tools as built–in functions of information system artifacts. The authors state that such systems could even adjust to the affective state of the individual using it. Therefore, the design science research on information system can build on advances in the fields of affective computing (Nacke et al., 2011).

A need arose to combine psychophysiology modalities, to enrich the interaction experience, and allow further information to be extracted from the data. However, this proposition carries with itself a whole new set of problems. This *multimodal interaction* approach combines ECG, GSR, *electromyography* (EMG), *electroencephalography* (EEG), pupil diameter and many other modalities, now sharing the same relevance of making interaction richer and more immersive. Such interaction communicates emotions, or rather emotional states, to the computer to recognize, interpret and communicate emotions back, in an attempt to create believable artificial intelligence (Picard, 2000). The author further states that it is possible to make the user interaction more natural by recognizing emotional states of the users, exhibiting sympathy and exhibiting emotional intelligence, which is a trait of intelligent beings.

The emerging discipline of affective computing, as founded by Picard (2000), informs about the integration of emotion into game development research (Hudlicka, 2008). Game development is affective if it can integrate players' emotions, while those players become consciously aware of how their physiology controls the game; as well as, when they are able to consciously control their reactions, in a biofeedback loop mechanism (Gilleade et al., 2005). The authors further state that developing affective games becomes an integration of player' physiological responses in influencing gameplay.

2.5 Brain–Computer Interface

Brain–Computer Interface (BCI) methodology uses EEG to extract signals providing activation of different brain regions and frequency of the brain rhythms. BCI is the neuroscientific interaction technology in the field of *human–computer interaction* (HCI) (van Erp et al., 2011). According to Schalk et al. (2008), BCI has four defined criteria: (i) the recording device must rely on signals recorded from the brain; (ii) there must be at least one recordable brain signal that the user can learn how to manipulate intentionally; (iii) real–time processing must be available; and (iv) the user must obtain feedback. Note that this does not exclude systems that utilize additional input (e.g., mouse or keyboard), as well as, recorded data from the brain. BCI methods hold a noticeably reduced response time compared to the other physiology sensors, with the complex data analysis tradeoff. The concept of non–invasive BCI is a relatively new field of research, and it is not until recent years that equipment has made it possible to use on–line signals produced by the human brain as a vehicle for HCI. The definition of non–invasive BCI states that there is no other discomfort to the user than that of wearing the sensor; in other words, there is no injury to the human body involved.

Technique	Accuracy
K–Nearest Neighbor	70.37%
Regression Tree	62.96%
Bayesian Network	59.26%
Support Vector Machine	77.78%
Artificial Neural Networks	70.37%
Random guess	33.33%

Table 2.1: Different classifiers of emotion and their maximum classification accuracy based on three datasets (Sohaib et al., 2013).

For wide acceptance of the mentioned technology into game development, usability and player experience will need to be considered when designing such systems (Bos et al., 2010). It would be helpful if Bci could inform games about an emotional state of their players (e.g., if players are not in the flow envisaged by the game designer) (Nijholt et al., 2009). Authors further state that measuring player experience and affect for adaptation of game development to provide intended experience is an important issue, but probably even more interesting are games developed to allow control from the brain activity that is consciously produced by players. The goal of evaluating Bci games is the achievement of the design choices that were made in the game, the functionality of the game, the ease with which the user could learn and memorize the control of the Bci, and with what accuracy they could control the game. In other words, the achievement can be seen as everything the user experienced during the game (Bos et al., 2010).

Bci applications emphasized the need of the emotion recognition using Eeg signals. Systems that can make interpretations about affective states based upon physiological data are of particular interest. Different classifiers of emotion and their maximum classification accuracy can be seen in Table 2.1. Linear classifiers are considered to be the most appropriate classification technology due to their simplicity, speed and interpretability (Fisher, 1936; Efron and Hastie, 1997; Jaakkola and Jordan, 1997). However, non–linear classifiers are considered to be the most appropriate when it comes to signal features and cognitive states (Wilson et al., 2010; Millán et al., 2004). From a number of machine learning algorithms previously used for classifying Eeg data, common ones associated with affective/emotional states are K–Nearest Neighbor, Regression Tree, Bayesian Network, Support Vector Machine and Artificial Neural Network (Rani et al., 2006).

2.6 Biofeedback

A core concept in affective computing is biofeedback. Biofeedback aims at displaying physiological parameters such as HR or GSR visually or acoustically (Nacke et al., 2011; Ouwerkerk et al., 2008; Lehrer et al., 2000). By this illustration, the participants get direct feedback on their visceral processes, processes which are for most participants usually almost unperceivable (Dawson et al., 2011; Crone et al., 2004). Biofeedback can increase the users' attention to their emotional states, which in turn can improve *interoception* (i.e., the conscious awareness of one's own physiological processes) (Wang et al., 2010; Critchley et al., 2004). Biofeedback was found to help successfully reduce heart diseases, stress, and anxiety (Varvogli and Darviri, 2011).

While contingent reinforcement in biofeedback has been a part of the biofeedback investigations from the early start, it has not been thoroughly investigated (Bergman and Johnson, 1972). Typical reinforcement commonly used in the early investigations of biofeedback–assisted HR control has included monetary bonuses (Clemens and Shattock, 1979), raffle tickets (Choi and Steptoe, 1982), points on a clock–type counter (McCanne, 1983), verbal praise (Riley and Furedy, 1981), and the opportunity to view commercial television programs (Scott et al., 1973) or attractive images (Schwartz, 1972). The specific contingencies used have varied across the early investigations as well, with non–contingent compensation (Perski and Engel, 1980), reinforcing successful HR control (Reeves and Shapiro, 1982), reinforcing improved performance on an accompanying behavioral stressor task (Malcuit and Beaudry, 1980) and reinforcing both improved task performance and successful HR control (Larkin et al., 1990).

There are three biofeedback presentation methods in serious games (Parnandi and Gutierrez-Osuna, 2017): a) visual biofeedback presents physiological information directly to the user (e.g., via a visual display); b) gameplay biofeedback presents the physiological information indirectly through subtle changes in gameplay (e.g., by changing game difficulty in proportion to the players' stress levels); c) combined biofeedback, delivers visual biofeedback and gameplay biofeedback simultaneously. For example, tunnel vision mimicking extreme stress has been modeled using visual (direct) biofeedback presentation which was connected to arousal (Grossman and Christensen, 2007). Such presentation in the serious game obstructed the display with a red texture that partially covered the field of view, limited only to when a small oval portion of the center remained visible to interfere with the task. This required corrective actions from the user, but the obstruction did not completely cover the view in order to avoid obstructing the action and infer a sense of loss of control (Bouchard et al., 2012). In the other examples, blur has been administered in the other study (Hilborn et al., 2013), and arousal in the form of an arousal meter (Caria et al., 2007) has been used in the Philips Rationalizer, which was designed to make the arousal levels 'intuitively clear' (Djajadiningrat et al., 2009). Moreover, audio (direct) biofeedback presentation was modeled using the sound of pre–recorded heartbeats increasing in frequency and loudness, which accompanied the visual feedback, and both physiological parameters were integrated to provide biofeedback of arousal levels, by referring to a baseline recorded while the serious game was loading. Gameplay (indirect) biofeedback presentation has been administered through the difficulty of the serious game task which was connected to arousal. Such arousal values are displayed visually in serious games, or they alter the gameplay (difficulty) (Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012).

The operant conditioning of Hr through biofeedback consists of two interrelated components: provision of Hr information to the individual; and the associated reinforcement presented in various forms to the individual following successful regulation of cardiac activity (Bergman and Johnson, 1972). Previous investigations using control groups (one of these groups was instructed to reduce Hr, and the other was not) examined the effect of a score contingent reinforcement in conjunction with Hr biofeedback which used both visual (direct) and gameplay (indirect) biofeedback presentation to make comparable statements of both Hr control and game performance (Larkin et al., 1990). The authors found that the combined (visual and gameplay) biofeedback presentation significantly reduced their Hr responses on a serious game task in comparison to all other groups. While these findings suggested that contingent reinforcement presented with Hr biofeedback may facilitate the skill–acquisition of down–regulation of Hr reactions under stress, they failed to elucidate what impact the operant conditioning would have upon down–regulation of Hr reactions to the task, independent of Hr biofeedback.

Operant (instrumental) conditioning is the central mechanism in the acquisition of skills using biofeedback in serious games (Parnandi and Gutierrez-Osuna, 2017). More specifically, the concept of negative–reinforcement instrumental conditioning, where the target instrumental conditioning behavior eliminates the occurrence of an aversive stimulus which leads to the reinforcement of that behavior (skill–acquisition). Biofeedback in the negative–reinforcement instrumental conditioning context forces an individual to down–regulate their arousal level (i.e., the instrumental response) to reduce game penalty (the aversive outcome) and progress in the serious game (Cannon-Bowers and Salas, 1998). In other words, there is a negative contingency between the instrumental response and aversive outcome. Furthermore, negative–reinforcement instrumental conditioning increases the likelihood that instrumental behavior will be repeated in the future indicating skill–transfer (Domjan, 2014). Concerning these mechanisms, an issue emerged regarding generalization and skill–transfer of biofeedback–assisted reductions in Hr reactivity to tasks resembling naturally–occurring stressors, during which biofeedback is not available (Larkin et al., 1992). While most investigations of Hr biofeedback have used validation tasks to assess skill–transfer of individuals' abilities to down–regulate Hr in the absence of biofeedback (Larkin et al., 1990), none have assessed the generalization of skills acquired through training to stressful tasks not encountered during training.

Although commonly used, only a few investigations have systematically examined the role of reinforcement contingencies in biofeedback–assisted changes in Hr. The investigation of the effect of incentive upon feedback–assisted Hr acceleration and deceleration found that participants in an incentive group (monetary reward) showed better Hr control than those in a non–incentive group (no reward) (Lang and Twentyman, 1976), which has been confirmed to warrant a more empirical investigation of the effect of various incentive conditions (Johnston and Lethem, 1981).

The somatic hypothesis and research in psychophysiology found that the experience of both emotions and stress are known to be accompanied by a physiological state of arousal (Lehrer and Vaschillo, 2004), which manifests itself through changes in physiology (Grandey, 2000). *Heart–rate variability* (Hrv) is quantified by measuring the interval variability between successive heartbeats (Rr intervals) in the Ecg signal (Kim et al., 2013). Evidence shows that Hrv is associated with emotions; therefore if this psychophysiological measure is incorporated in biofeedback self–regulation practice, it can improve emotion–regulation and cognitive functioning as suggested by the somatic hypothesis (Lehrer and Vaschillo, 2004).

Control conditions have been a part of the biofeedback investigations from the early start, where minimal cardiac information (e.g., instructions to alter Hr unaccompanied by biofeedback) has been compared with groups which received enhanced cardiac information (e.g., feedback accompanied by an auditory Hr signal) (Larkin et al., 1992). In addition, as individuals are rarely focused on the physiological and psychological effects of stress during a demanding task, they would benefit from the continuous information of arousal while engaged on a highly stressful serious game task; therefore, they could acquire the skill to detect signs of stress and use an appropriate coping strategy (Jerčić et al., 2012). Moreover, biofeedback in serious games reinforces the skill–acquisition process with a positive reward and an increase in perceived self–efficacy when regulation skills are mastered. This is achieved by tailoring the serious game session for each individual engaged in mastering the regulation skills, in order to motivate better performance by negative reinforcement through serious game difficulty based on real–time physiological data (player–centered design).

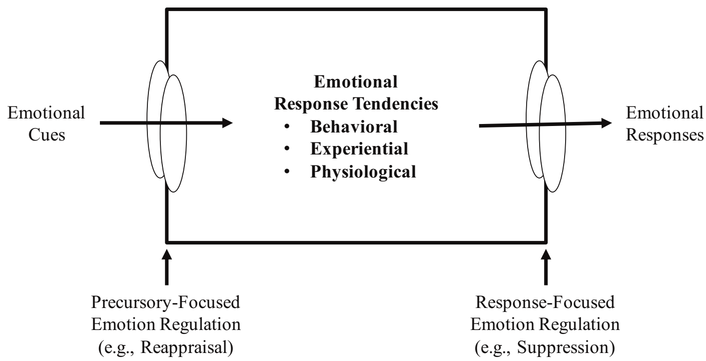

Figure 2.2: A consensual process model of emotion that highlights two major classes of emotion–regulation. Adapted from Gross (1998b).

2.7 Emotion–Regulation

Emotion–regulation can be described by the process model of Gross (1998b) which is widely known and acknowledged in the field of emotion–regulation strategies. It relies on the assumption that emotions are generated in an emotion–generative process. A broad distinction which the author draws is the one between precursory–focused and response–focused emotion–regulation strategies, as depicted in Figure 2.2. Precursory emotion–regulation strategies apply while the emotion is still unfolding and has not reached its peak. Regulation is influenced by two broad categories of effects: on–task attentional interest and off–task attentional distractors. On–task interest is stronger if the task is important, intrinsically motivating, or rewarded for completion. Off–task distractors are things that make it harder to maintain focus on the desired task, such as disturbing noises, interruption from people, and distractions of an affective character such as sadness or anger. People tend to use one of two main, broadly defined, strategies to deal with emotions emerging when facing difficult and stressful tasks (Wallace et al., 2009). These strategies are *cognitive reappraisal* and *suppression*. Gross (1998b) defines cognitive reappraisal and suppression as the following. Cognitive reappraisal is a type of cognitive change defined as "construing a potentially emotion–eliciting situation in non–emotional terms", it is the process of constructing a personal interpretation of a situation. Expressive suppression (henceforth referred to as suppression), is a type of response modulation defined as "inhibiting ongoing emotion–expressive behavior", it is the process of deliberately trying to stop thinking about certain thoughts. Response–focused 'suppression' strategy operates on the emotion that has already completely unfolded, where it alters and controls the experiential, behavioral, and physiological response. Precursory 'cognitive reappraisal' strategy applies while the emotion is still unfolding and has not reached its peak (Wallace et al., 2009). Reappraisal generally reduces emotional experience and behavior expression; furthermore, it has no impact on memory. On the other hand, suppression generally reduces behavior expression, but the emotional experience is unaffected, and memory is impaired (Fenton-O'Creevy et al., 2012b).

Fenton-O'Creevy et al. (2012b) found that reappraisal tendencies correlated with better performance, while suppression tendencies correlated with worse performance. They explained their results by stating that the cognitive load was higher for suppression than for reappraisal, since reappraisal happens before the onset of an emotion, rendering the emotional response lower and thus creating a less off–task attentional distraction. In line with the description in the previous paragraph, suppressors tend to continually push down emotions, ignoring the fact that they exist and are continuously affecting them. Therefore, the degree of inhibitory control is negatively correlated with susceptibility to framing effects in a risky–choice task (Sütterlin et al., 2011). On the other hand, reappraisers tend to re–evaluate situations positively (Wallace et al., 2009). Both emotion–regulation strategies take up cognitive resources (Wallace et al., 2009). However, the authors also state that suppressing emotions generally takes up more cognitive resources in comparison to the reappraisal strategy when encountering undesired emotions. Hence, emotional suppression can eventually take up so many cognitive resources that it can reduce one's performance in decision tasks compared to the strategy of emotional reappraisal. While it is possible to induce an intentional cognitive reappraisal strategy, Sokol-Hessner et al. (2009) showed that applying this strategy can mitigate physiological responses to losses relative to gains and also reduce loss aversion. Evidence shows that the emotion–regulation skill can reduce aversion of negative emotions and contribute to advantageous decision–making by helping to mitigate emotion–related decision biases (Fenton-O'Creevy et al., 2011; Heilman et al., 2010; Gross, 2009). Individuals with low skills are prone to make disadvantageous decisions with undesirable consequences (Sütterlin et al., 2011). Research shows that the effects of practicing emotion–regulation through biofeedback on the decision–making performance in the context of serious games, allow individuals to practice the skills of down–regulating high level of stress and arousal, to perform better on a decision–making task (Parnandi and Gutierrez-Osuna, 2017; Astor et al., 2013b; Hilborn et al., 2013; Kim et al., 2013; Bouchard et al., 2012; Jerčić et al., 2012). However, research in cognitive neuroscience and psychophysiology found that emotion–regulation has an influence on the underlying visceral processes of emotional experience (Martin and Delgado, 2011). Individuals with brain injury usually have a significant deficit in self–regulation (Bechara, 2000). In the field, it has been shown that traders and investors with high emotion–regulation skills perform better in trading (Fenton-O'Creevy et al., 2011); moreover, soldiers who are frequently exposed to traumatic events and acute stressors could benefit from fostering better emotional resiliency since it might improve their performance in decision–making (Parnandi and Gutierrez-Osuna, 2017).

Emotions and their regulation can be both consciously and subconsciously processed (Gross, 2009; Williams and Bargh, 2009), where increased emotional awareness is crucial for improving it (Williams and Bargh, 2009). The steps to correct treatment have been validated in the previous reviews (Culbert, 2017) and they equal to: 1) awareness to objectively discriminate the mind–body differences between sympathetic nervous system arousal (stress) and parasympathetic nervous system dominant states (relaxation response); 2) consistent regulation of a given psychological or physiological modality towards a target direction (first in a controlled environment); 3) applying this method to regulate in the appropriate life situations, 'in the moment' (skill–transfer); 4) change to restructure Cns and rebalance (baseline) Ans with regular long–term daily practice to an emotionally resilient pattern. As argued, the first step towards emotion–regulation is to make individuals aware of those emerging emotions (Culbert, 2017).

Following the concept of emotion–regulation, emotions can determine and alter the way people perceive a particular situation and thereby influence the way they react to it (Astor et al., 2013b). Thus, emotions act as response tendencies: they suggest distinctive responses to a specific situation, which an individual may or may not follow (Gross, 1998b). In this sense, emotions reflect the ongoing adjustments to continually changing environmental demands (Thayer et al., 2009). Evidence shows a significant relationship between effective emotion–regulation, better awareness of own current emotional state and also a sophisticated skill in down–regulating high levels of arousal (Fenton-O'Creevy et al., 2012b). Therefore, increased emotional awareness is crucial for enhanced emotion–regulation. While important decisions generally elicit strong emotions in decision makers, there are strong interpersonal differences in the capabilities of adequate emotional processing (Lo et al., 2005; Lo and Repin, 2002; Loewenstein, 2000). The applied emotion–regulation strategies seem to be determined by personality traits and individual psychological capacity (Gross, 1998a). However, it is important to note that most of these studies refer to contextualized situations and "the context can influence how we process stimuli that may have affective properties" (Rolls and Grabenhorst, 2008, p. 230). In other words, affective processes are context dependent (Astor et al., 2013b).

2.8 Decision–Making

There is extensive evidence that emotions are a key factor that critically influences human decision–making (Adam and Gamer, 2011; Loewenstein, 2000). It can also be stated that the original division between emotions and cognition drawn in the past is rather unrealistic (Phelps, 2006). The traditional consequentialist perspective idealizes decision makers as a perfectly rational cognitive machine which favors utility maximization when confronted with decisions (Rasmusen, 2006). Therefore, it does not take into account the influence of emotions on cognition and decision–making, and cannot explain ample evidence that their assumption does not hold for most human decision makers (Bechara and Damasio, 2005; Shiv et al., 2005; Sanfey et al., 2003). Evidence shows that arousal can significantly threaten rational cognition and decision–making (Kuhnen and Knutson, 2005). Poor decision–making performance is often associated with poor impulse control strategies, high emotional reactivity, and exposure to high levels of arousal (Peterson, 2007). Increased awareness of emotional states can help individuals to identify a state of high arousal, which might impair with their decision–making performance (Fenton-O'Creevy et al., 2012a; Lo et al., 2005). It has been found that individuals aware of their emotional states avoid being overpowered by their emotions, and need to develop skills to interpret and down–regulate those affective states of high arousal (Fenton-O'Creevy et al., 2012a; Peterson, 2007). This can result in enhanced decision–making performance (Biais et al., 2005). Currently, psychology sees decision–making as a dual–process framework which works in two fundamentally different ways depending on the context. The slow and analytical rational system, and the fast and intuitive experiential system (Kahneman and Frederick, 2002). This framework is especially relevant in contexts where fast processing is necessary. In such contexts, humans tend to follow an 'affect heuristic' and rely on their emotions to help evaluate situations instantly, or to process informational overload when they have to come to quick decisions (Jerčić et al., 2012; Slovic et al., 2007). The authors state that this may require a delicate balance of activation of one dual–process framework mechanism or the other, which may lead to reduced performance (Kuhnen and Knutson, 2005). Cognitive neuroscience has given evidence that activation in specific neural circuits promotes and avoids varying types of risky financial mistakes, correlating with positive as well as negative affective states (Peterson, 2007; Kuhnen and Knutson, 2005). The affect heuristic proposes that decision makers integrate context–specific affective feelings into perception and their decision–making process. Furthermore, the context of the task has been identified as an important factor of how affective stimuli are processed (Rolls and Grabenhorst, 2008).

Bechara (2004) specifically accounted for the interplay of emotions and information processing, where emotions can contain relevant information. In their somatic marker hypothesis, Bechara and Damasio (2005) state that emotions may be both a source of biases and an important mechanism for optimal decision–making, where somatic markers (essentially emotional responses to information events) beneficially guide one's focus of attention. The authors showed that the experience of emotions is a necessary prerequisite for optimal decision–making. More specifically, the authors found that those individuals who performed worse on a task under ambiguity (compared to the healthy individuals) had brain lesions which are critical for the processing of emotions (Bechara and Damasio, 2005; Bechara et al., 1997). According to the theory, decisions are aided by emotions, in the form of elicited bodily states, during the consideration of future consequences and contemplation of different options for advantageous or disadvantageous behavior (Naqvi et al., 2006). Therefore, it has been found that emotions may have positive effects on decision–making performance (Seo and Barrett, 2007). However, it was shown that such emotional processes could also have disruptive effects in other contexts (Shiv et al., 2005). There is hence a bilateral effect of emotions: on the one hand, they are bias–inducing and hence malicious to the decision maker, but on the other, they also provide valuable knowledge in representing for example experiences one has gained in the past (Astor et al., 2011). Therefore, emotions cannot always "be trusted as leading to good or bad decisions" (Shiv et al., 2005, p. 438). Still, awareness of emotions and a well–functioning recognition of these bodily states is an essential prerequisite for deciding whether it is better to follow an emotional response or to inhibit it (Gross, 1998b).

Drawing from economic research there is ample evidence that economic decision–making can be biased to a considerable extent by levels of high emotionality and arousal (Adam et al., 2011; Loewenstein, 2000). Fenton-O'Creevy et al. (2011) detected a strong correlation between traders' ability to regulate emotions and their financial performance. The authors found that a better perception and awareness of emotional states was associated with high performing traders. Most interestingly, these traders are also more advanced in regulating their emotions. Further laboratory and field studies showed that emotional processes can also have disruptive effects and that they cannot be trusted as leading to optimal decisions (Shiv et al., 2005; Bechara and Damasio, 2005; Shefrin and Statman, 1985). These disruptive effects include the disposition effect (Shefrin and Statman, 1985), loss aversion (Sütterlin et al., 2011; Sokol-Hessner et al., 2009), and the phenomenon of auction fever (Adam et al., 2011). In the field, traders and investors already realized that emotions and decision–making biases are even beyond conscious awareness, and are therefore highly interested in regulating such emotional dispositions (Fenton-O'Creevy et al., 2011).

2.9 Human–Robot Interaction

Hri is a field of study dedicated to understanding, designing and evaluating robotic systems for use by or with humans (Goodrich and Schultz, 2007). Authors further define that, such interaction requires communication between robots and humans. It is the process of working together to accomplish a goal. Such communication may take several forms, one of which is *proximate interaction*, where humans and robots are collocated (i.e., a collaborative robot is in the same room as humans). Proximate interaction with mobile robots may take the form of a robot assistant and may include physical interaction. Mobility with apparent intention in physical space is likely for a robot to be perceived as animate, eliciting social responses (Scholl and Tremoulet, 2000). Such *social interaction* where humans and robots interact as peers or companions appears to be proximate rather than remote; moreover, it includes social, emotional, and cognitive aspects.

Hri challenges include supporting effective social interactions through cognitive and emotional cues, as well as, through natural interactions such as gestures (Goodrich and Schultz, 2007). To make the exchange between humans and robots beneficial in some sense researchers attempt to understand and shape the interaction itself (e.g., training and performance on a serious game task). Such investigation seeks to improve the quality of Hri by developing serious games that will rely on social and emotional cues and therefore be more natural, intuitive and familiar for users (Breazeal, 2002). Traditional human–computer interaction approaches and methods become less applicable as intelligent autonomous robots are perceived to be more anthropomorphic than other computing systems (Kiesler and Hinds, 2004). Such social interaction aims to bring interaction closer to the physical context of humans.

The individuals who are merely presented with the information are more passive in contrast to the ones who are interacting with it (Dror, 2008). Moreover, down–regulation of high levels of arousal may only be practiced in an arousing environment. As outlined by Grandey (2000, p. 99), "the experience of both emotions and stress are known to be accompanied by a physiological state of arousal". Serious games that use robot collaborators might elicit strong emotions in players, and they can be an engaging environment which can elicit arousal. Previous research employing questionnaires has shown that embodied robots are sometimes as engaging as humans (Jung and Lee, 2004; Burgoon and Bonito, 2000). Although there is certain popularity of electronic games in current research methods, many traditional games are played in the physical world and require tangible interaction (Xin and Sharlin, 2007). The physicality of games is important because humans perceive robots as physical entities, capable of interaction within the physical world while having access to the virtual domain, much like a traditional computer game system. Such Hri–enhanced traditional serious games can support this physical and virtual duality.

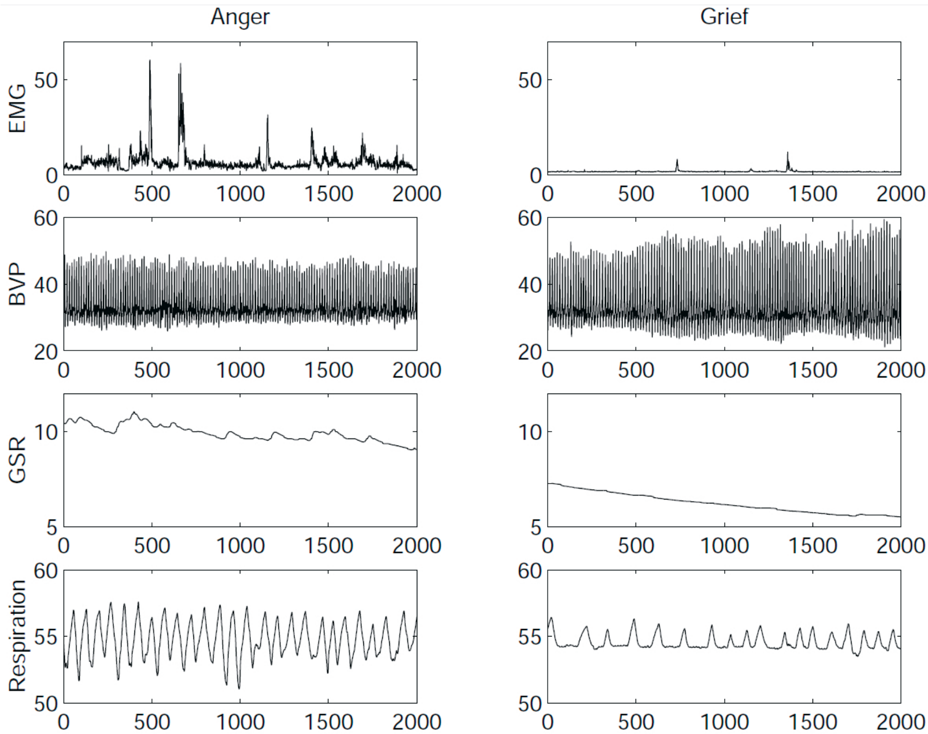

Figure 2.3: Example of four physiological signals measured from an actress while she intentionally expressed anger (left) and grief (right). From top to bottom: Electromyogram (microvolts), blood volume pressure (percent reflectance), Gsr (microSiemens), and respiration (percent maximum expansion).The signals were sampled at 20 samples a second. Each box shows 100 seconds of response. The segment shown here are visibly different for the two emotions, which was not true in general. Reproduced from Vyzas and Picard (1998) with permission.

2.10 Technological Implementation

Contemporary psychophysiology constitutes of many physiological modalities (see Figure 2.3), out of which this thesis is focusing on *electroencephalography* (EEG); further supported by measures of *skin resistance* (GSR), *cardiovascular measures* (HR; HRV), *muscle activity* (EMG), and *pupillometry* changes in *pupil diameter* (PD) and in regards to cognition and emotion. These measures may provide key information regarding the intensity and quality of an individual's internal experience and affect. Furthermore, they can easily be digitized and may eventually be unobtrusively (non–invasively) monitored, making them very accessible to pattern recognition techniques (Scheirer et al., 2002). The above–mentioned physiological modalities have been detailed in the following sections, while their motivation and application in serious games has been given in the next chapter.

2.10.1 Electroencephalography

EEG is the recording of electrical activity along the scalp, indicating cognitive/conscious states of attention, alertness, and drowsiness. Methods like *evoked potentials* and *event–related potentials* average EEG responses which are time–locked to the presentation of a stimulus of some sort (e.g., visual, somatosensory, or auditory). They offer us the possibility of analyzing EEG responses. The *event–related de/synchronization* is the power of certain frequencies in the brain, and it gives information on the affect a person is feeling. The interpretation has largely been framed in terms of specific frequency bands associated with different cognitive and emotional states (Andreassi, 2007): alpha waves (8–13 Hz) with relaxation, reflection and inhibition; beta waves (14–30 Hz) with alertness, anxiety, concentration, mental and physical activity; delta waves (0.5–3.5 Hz) with deep sleep; theta waves (4–7 Hz) with drowsiness, pleasure, displeasure, idling and inhibition; kappa waves (10 Hz) with thinking; and gamma waves (resting frequency around 40 Hz, modulated by 3–5 Hz) with cross–modal perception and perceptual recognition. Regardless of the analysis goal, most of the signal processing methods and classification algorithms for EEG have been developed in the context of building BCIs (Bashashati et al., 2007; Lotte and Congedo, 2007), and researchers are constantly seeking ways for developing new approaches for recognizing affective states from EEG and other physiological signals. Evidence suggests that stronger alpha waves in the right frontal hemisphere are associated with withdrawal (negative emotions), while those in the left frontal hemisphere are associated with approach (positive emotions) (Larsen et al., 2008). These techniques are used in cognitive science, cognitive psychology, and psychophysiological research. A BCI is a direct communication pathway between the brain and an external device, a system which allows someone to communicate information about their mental state without the use of the peripheral nervous system (Calvo and D'Mello, 2010). The asymmetry among left and right brain hemispheres are the major areas where the emotion signals can be captured (Larsen et al., 2008). According to a model developed by (Davidson, 2003), the two core dimensions – arousal and valence – are related to the asymmetric behavior of emotions in EEG.

2.10.2 Galvanic Skin Response

GSR is a method of measuring the electrical conductance of the skin, where the skin momentarily becomes a better conductor of electricity when either external or internal physiologically arousing stimuli occur and excite the eccrine sweat glands under the skin, controlled by the SNS (Dawson et al., 2011; Burgoon and Bonito, 2002). GSR includes short–term phasic responses to specific stimuli, and relatively stable, longer–term tonic levels (Andreassi, 2007). Measurements are typically taken from the palm, the fingers, or the soles of the feet, where eccrine sweat glands are most densely distributed, providing the strongest signal variations. Since there is a relationship between sympathetic activity and emotional arousal, GSR has been closely linked to both emotion and attention (Scheirer et al., 2002), although it is used as an indication of psychological or physiological arousal. There are numerous accounts reported of a linear correlation between GSR and arousal (Braithwaite and Watson, 2013; Sequeira et al., 2009; Dawson et al., 2007; Lang et al., 2005; Haag et al., 2004; Stern et al., 2001; Lang et al., 1993; Plutchik, 1984). Even though GSR has been found to indicate physiological valence, the correlations were not always as strong, and greater confidence in interpretations has been obtained by combining GSR with other measures, such as ECG (Andreassi, 2007).

Using GSR data to measure arousal in a continuous stimulus setting requires at least three steps in data processing and analysis (Leiner et al., 2012). The **first step** is data cleaning and decomposition of the signal into its tonic and phasic components. Tonic skin conductance refers to an ongoing or baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event–related changes that are caused by a momentary increase in skin conductance (resembling a peak superimposed on tonic skin conductance). The *skin conductance level* (SCL) is typically approximated by frequency filtering, statistical modeling or simple linear interpolation between the skin conductance measures that are not overlaid by responses (Kettunen et al., 1998). The **second step** is parameterization, and for the phasic parameters, this process includes massive abstraction of the phasic signal component, e.g. *non–specific skin conductance responses* (Ns–scrs). Changes in arousal within periods shorter than two minutes are likely not to be indicated using the Scl (Vossel and Zimmer, 1990). This problem is particularly limiting in trials shorter than two minutes. When the Scl's temporal precision is insufficient, the rapidly reacting phasic changes observed as Ns–scrs appear to be a more promising focus: their number during a given period is a prominent phasic–based indicator of arousal (Boucsein, 2012). Significant increases in the frequency of Ns–scrs are also commonly seen in high arousal situations and as such can be viewed as an indicator of background arousal. Arousal measured through Ns–scrs has approximately 40–60% variance due to internal or external disruptions. Boucsein (2012) states that a typical rate of 1–5 fluctuations per minute is expected at rest (this number can be up to 20 under high arousal).

In the **third step**, it is necessary to consider that arousal is indicated by the frequency of spontaneous responses and by the amplitude of the (significant) *skin conductance responses* (Scr), where greater importance is attributed to Scrs with larger amplitude (Bach et al., 2010). It is essential to use a fixed time–period of measurement when analyzing the frequency of peaks and their amplitudes across conditions and participants (Boucsein, 2012). Analysis of amplitudes of Ns–scrs and their standard deviation could also provide additional indicators of tonic arousal (Braithwaite and Watson, 2013). However, any function of Scr number and amplitude remains highly sensitive to the threshold that distinguishes an Scr from noise. The (arbitrary) de–facto standard threshold is a minimum amplitude of 0.05 muS (Boucsein, 2012; Dawson et al., 2007). An additional step includes minor conductance fluctuations that could be spontaneous fluctuations or noise. Such noise may be a consequence of instruments' quality, the environment, or interpersonal differences. Interpersonal differences in Gsr are substantial, and neither the instrument nor the environment is likely to change during one test (Ben-Shakhar, 1985). Therefore, the noise will not decrease the accuracy of the results obtained by comparing different timeframes in the same session. The prospect of misinterpreting noise as skin responses is a minor drawback given the advantages of removing a possible source of systematic error.

Gsr is also known to be influenced by brain structures such as the hypothalamus, the limbic system and frontal cortical areas (Delplanque and Grandjean, 2008). Since these areas have been associated with emotional processing, Gsr can be considered as a window on emotional brain activity (Chanel, 2009). Moreover, skin conductance is not controlled by the Pns (Bach et al., 2010). In contrast, Hr is a measure that reflects the activity of both the sympathetic and parasympathetic branches of the Ans. Furthermore, sympathetic and parasympathetic activation is almost instantly reflected in changes in Hr (Berntson et al., 2007).

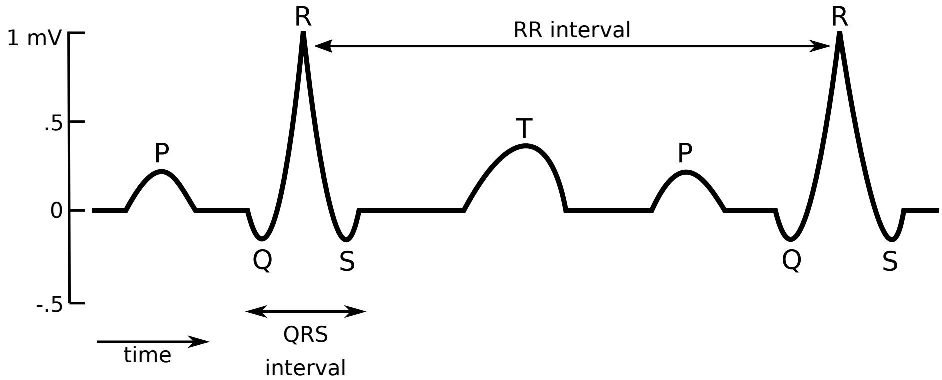

Figure 2.4: An example of a decomposition of the ECG signal onto the characteristic reflections referred to as P, Q, R, S, and T waves. The time between the most characteristic R peaks in the signal is called RR interval and it presents a base on which HR and HRV are computed. Adapted from Hamilton (2002).

2.10.3 Electrocardiography

ECG measures the rate and regularity of heartbeats by detecting the peaks of the highly positive R–waves in the signal (Houghton and Gray, 2014). Numerous studies have found that ECG is a good indicator of physiological valence as strong correlation has been found (Xu et al., 2010; Andreassi, 2007; Cacioppo et al., 2000; Papillo and Shapiro, 1990; Malmstrom et al., 1965). HR represents the number of heartbeats per unit of time, under the influence of SNS and the PNS. Physiology sensors measuring HR are beneficial since they are highly unlikely to fall off or be affected by movement since various HR monitors are designed on the same principles (e.g., waistband and wireless connection) that are used in sports (Giles et al., 2016). The arousal detection method was employed to make the inferences from the ECG signal, which decomposes the signal into the characteristic reflections (referred to as P, Q, R, S, and T waves) and detects the most characteristic R peaks in the signal based on which HR is computed (Hamilton, 2002), as illustrated in Figure 2.4. Therefore, this method is named the QRS detection method. If peaks are measured precisely enough, they can be used to extract HRV which represents variability between heartbeats intervals. In the field of psychophysiology, there is great interest in HRV since it is related to emotional arousal/valence (Scheirer et al., 2002; Rowe et al., 1998). Two distinct measures of HRV are *the standard deviation of normal–to–normal heartbeat intervals in time domain* (SDNN) and *the ratio of low– and high–frequency powers after the fast Fourier transform* (LF/HF) (Acharya et al., 2006). Wang and

Huang (2014) state that these two measures are employed as two dimensions in physiological valence/arousal model. HRV is highly correlated with emotion (Lane et al., 2009; Rainville et al., 2006; Appelhans and Luecken, 2006; Gendolla and Krüsken, 2002; McCraty et al., 1995), while some evidence reveals that SDNN is a valid indicator of valence in the physiological perspective (Kemp et al., 2011; Muller and Ellis, 2011; Geisler et al., 2010; Licht, 2008; Koelsch and Remppis, 2007; Carney and Saunders, 1995). SDNN presents the variation of the circulatory system and LF/HF presents the balance of the SNS and PNS (American, 1996). Xu et al. (2010) showed over 90% detection accuracy of both positive and negative valence for multiple participants based on ECG recordings. The on–line computation of arousal values available in serious games has also been based on the physiological measure of HR and HRV (Astor et al., 2013b).

For shorter recordings of 2 to 5 min, three main spectral components are distinguished in a spectrum (Malliani et al., 1991; Pagani et al., 1986; Akselrod et al., 1981; Hirsch and Bishop, 1981; Sayers, 1973): very low frequency (VLF) – the power related with the frequency band lower than 0.04 Hz; LF – the power related with the sympathetic nervous system activity of the heart with the frequency band between 0.04–0.15 Hz; and HF – the power associated with the PNS activity of the heart (in the frequency band 0.12 or 0.15–0.4 Hz) components.

Small deviations from a normal heartbeat are named ectopic heartbeats. Ectopic and arrhythmic beats in the ECG recording may alter the estimation of HRV by introducing extra or skipped heartbeats in the analysis, as well as missing and noisy data. During an arrhythmic event, the electrical activity of the heart is irregular or is faster or slower than usual. Correct interpolation or linear regression on preceding/successive heartbeats of HRV signal or its auto–correlation function may reduce this error. Preferentially, short–term recordings which are free of ectopy, missing data, and noise should be used. In the cases where only ectopic–free short–term recordings are acceptable, data interpolation should be used (Kamath and Fallen, 1995). The authors advise reducing the selection bias by considering the possibility that the data analysis may be influenced by ectopy in such cases. The relative number and relative duration of RR intervals which were omitted and interpolated should also be quoted.

The distribution of the power and central frequency on LF and HF may differ based on the heart rate changes influenced by the autonomic modulations (Malliani et al., 1991; Furlan et al., 1990; Pagani et al., 1986). The LF and HF components are the relative power of each component proportional to the total power reduced by the VLF component and they are measured in normalized units (Malliani et al., 1991; Pagani et al., 1986). The VLF component is still questionable since a specific physiological explanation regarding these heart rate changes has not been identified. Thus, VLF assessed from short–term recordings is a dubious measure and should be avoided. The representation of Lf and Hf in normalized units emphasizes the controlled and balanced behavior of Pns and Sns. Nevertheless, normalized units should always be quoted with absolute values of Lf and Hf power in order to completely describe the distribution of power in spectral components. Stressors are often associated with an increase in sympathetic cardiac control, a decrease in parasympathetic control, or both. Associated with these reactions is a frequently reported increase in Lf centered around 0.1 Hz power, a decrease in Hf, and/or an increase in the Lf/Hf ratio (Jain et al., 2001; Delaney and Brodie, 2000; Hughes and Stoney, 2000; Friedman et al., 1996; Berntson and Cacioppo, 1994). The parasympathetic influences are pervasive over the frequency range of the Hr power spectrum, whereas the sympathetic influences 'roll–off' at about 0.15 Hz (Saul, 1990). Therefore Hf in Hrv primarily reflects parasympathetic influences within the lower frequencies (< 0.15 Hz), having both an influence from Sns and Pns. The differential effects of the Ans on the heart, and thus the timing of the heartbeats, are a consequence of the differential effects of neurotransmitters for the Sns and Pns. Sympathetic effects are slow, measured in seconds, whereas parasympathetic effects are fast, measured in milliseconds. Therefore, parasympathetic influences are the only ones capable of producing rapid changes in the beat to beat timing of the heart. Typically, Hrv is quantified by measuring the interval between successive Rr intervals in the Ecg signal (Kim et al., 2013). The arousal computation is usually performed using the mean value over the last five heartbeat intervals since breathing affects Hr, where both have been correlated with arousal (Berntson et al., 2007).

Previous investigations gave evidence that diverse emotions evaluated through Hrv are correlated with the patterns of Ans activity (Demaree and Everhart, 2004; Levenson, 1992). Appelhans and Luecken (2006) state that Hrv is a measure of the continuous interplay between sympathetic and parasympathetic influences on Hr that yields information about Ans flexibility and thereby represents its capacity for regulated emotional response. It has been generally proposed that positive emotions are associated with parasympathetic cardiac control (Fredrickson, 2001; Porges, 1995), and to fast recovery from cardiovascular effects of negative emotions by reducing sympathetic activity (Fredrickson et al., 2000).

Previous investigations established a direct connection between the Cns and the Ans, which is reflected in Hrv (Thayer et al., 2009). While the and pre–frontal activity has been associated specifically with Hrv mediated by the vagus nerve (Lane et al., 2009). In the brain injury context, evidence shows that severe brain injury can cause dysregulation of the Ans (Galluzzi et al., 2009). Such individuals who suffer from autonomic dysfunction typically exhibit little modulation of Hr and low amplitude in the Hrv patterns (Tan et al., 2009). These symptoms are correlated with an individual's deficits in decision–making tasks that involve executive function (Thayer et al., 2009). The authors show evidence that Hrv is associated with certain specific executive functions: attention, the flexibility of behavior and control of emotions (Thayer et al., 2009). As mentioned earlier, breathing is also correlated with arousal in the form of *respiratory sinus–arrhythmia* (Rsa), where it refers to the component of the change in Rr intervals that is synchronized to the respiratory cycle (Lehrer et al., 2000). Moreover, Rsa may be a dominant component of the change in the RR interval when the individual's breathing is at an optimal frequency, which is referred to as 'resonant frequency' or 'coherence.' Previous investigations found that individuals with greater Hrv had significantly more correct responses in a working memory test and in a continuous performance test, as well as faster reaction time (Hansen et al., 2003). Furthermore, investigations of children with emotional and behavioral disorders found that higher scores on fast–paced decision–making tasks requiring executive control, the accuracy of response and inhibition of response in relation to changing information, were significantly associated with higher Rsa (Mezzacappa et al., 1998). These two physiological parameters, large amplitude modulation in Hrv and coherence or resonance of Rsa have been associated with executive functioning (McCraty et al., 2009; Thayer et al., 2009). Maximal regulation occurs at a particular cardiovascular 'resonant frequency' of the baroreflex system correlated to Hrv, typically 0.1 Hertz or a 10–second rhythm (Vaschillo et al., 2006; Lehrer et al., 2000), and it is thought to reflect a balance between the two branches (sympathetic and parasympathetic) of the Ans. Therefore, practicing Rsa through the Hrv biofeedback improves the skill–acquisition of emotion–regulation by influencing physiology (Moraveji et al., 2011; Mezzacappa et al., 1998). The goal of Hrv biofeedback is to help individuals increase the relative amount of Rsa in the Hrv signal (Kim et al., 2013). From both a psychological and physiological standpoint Rsa has been shown to be most closely associated with self–regulation , where the amplitude of Rsa is correlated with emotional disorders, emotional dysregulation, and inflexibility of behavior (Thayer et al., 2009; Karavidas et al., 2007).

2.10.4 Pupil

Pupillary response is the variation of the pupil size in response to a stimulus under the influence of Ans mediated by the optic and oculomotor cranial nerve. It is a physiological response that varies the size of the pupil, either resulting in constriction (Sns) or dilation (Pns), via activation or deactivation of the iris dilator muscles (Ellis, 1981). The pupil dilates in response to extreme emotional situations such as fear, or contacts in response to a sensory nerve activation, such

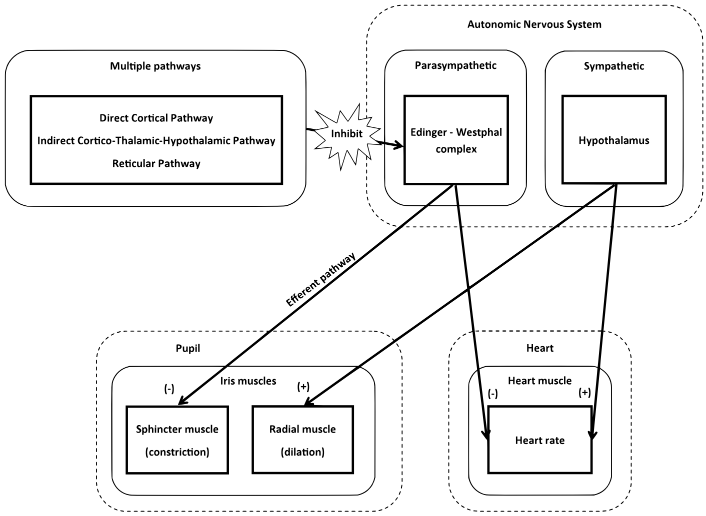

Figure 2.5: Model of the Ans control pathways for the pupil and the heart (Jerčić et al., 2018a).

as pain. Previous studies have found that pupil size variations are also related to both cognitive and affective cues, fear and pain being an example of pupil dilation response to stimuli. If emotions and pupil size variation were reliably associated with each other, then current eye–tracking technology would offer a possibility for unobtrusive monitoring of emotion–related reactions because there would be no need to attach sensors to users. In order to be able to evaluate the possibilities of using pupil size measurement for detecting the users' emotional responses, one needs to understand how emotions and pupil size variation relate to each other (Partala and Surakka, 2003). Eye tracking has also been used in conjunction with psychophysiological data and game metrics to evaluate the player experience (Zammitto et al., 2010).

It was previously established that the extent of interplay between Sns or Pns activities provide a measure of participants' emotion–regulation capabilities. As illustrated in the schematic view on Figure 2.5., those two synergistic pathways control pupil size and dynamics, operating on the smooth muscles of the pupil. The Pns pathway is mediated with the pathway originating in the Edinger–Westphal oculomotor complex in the midbrain and innervates the sphincter which is the circular muscle responsible for constriction, as seen in the reflex reaction to light. Inhibition of Edinger–Westphal complex results in relaxation of the sphincter muscles and, thus, significant dilation. The Sns pathway, mediated by the hypothalamus, innervates the radial dilator muscle of the iris responsible for dilation (Privitera et al., 2008; Steinhauer et al., 2004). Similarly, Berntson et al. (2007) states that the Hr is also a measure of activity of both the Sns and Pns part of Ans. Correspondingly, Sns activity tends to increase Hr, while Pns activity decreases it.

2.10.5 Electromyography

Emg is a technique for evaluating and recording the electrical activity produced by skeletal muscles. *Facial electromyography* (fEmg) refers to measurements of the muscles of the face indicating emotional reactions (Dimberg, 1990). Studies have found that the activity of the *corrugators supercilii* muscle (frowns) and the *zygomaticus major* muscle (smiling) are associated with emotional valence (Larsen et al., 2008). fEmg has been used as a technique to distinguish and track positive and negative emotional reactions to stimuli as they occur (Wolf et al., 2005).

Examples of contemporary use of Emg in game development research are: fEmg and Hr in addition to self–report ratings to index physiological arousal and emotional valence (Ravaja et al., 2006); a real–time emotional profile (i.e., flow and immersion) of gameplay by measuring Eeg, Hr, Emg, Gsr and using eye–tracking (Nacke and Lindley, 2010); and a case study on Emg, Gsr and Hr correlations with player gameplay experience in a first–person shooter game (Drachen et al., 2010).

2.11 Related Work

There exists a number of serious games in various fields of science, ranging from medicine, psychology, neuroscience, well–being, finance and many more. Detailed in this chapter are the serious games identified through the literature reviews performed in the research papers presented in Chapters 4–10, and through the systematic literature review presented in Chapter 11.

2.11.1 Serious Games in the Field of Healthcare and Well–Being

There are a number of serious games using biofeedback in the field of healthcare and well–being (Kato, 2010). These are mostly concerned with hyperactivity disorder, autism, substance abuse; and more specific targets such as pain, asthma, bladder control, or medical education on cancer (Wang et al., 2010). These games use various psychophysiological sensors to make the users aware of their medical states and provide a clear goal on how to improve them using feedback (Dunwell et al., 2010). Previous investigations gave evidence that biofeedback regarding physiological arousal using serious games can help people to manage their stress levels and relax to perform better in the games. These serious game artifacts include: Relax to Win (Sharry et al., 2003) where a game was created to help children who suffered from anxiety problems, where the level of relaxation has been measured using Ecg and Gsr; Froment et al. (2009) enabled participants to play the modified game of Unreal Tournament while being connected to Ecg, Gsr and body temperature sensors, to manipulate the effects such as screen blur and off–aim; Brainball (Hjelm, 2003) uses Bci to raise awareness of emotions in players.

2.11.2 Serious Games using Biofeedback for Emotion–Regulation Training

Psychophysiological signals can be seen as an objective measure of emotional states, as it is hard to manipulate them intentionally. Iancovici et al. (2011) gave an up–to–date overview on (serious) games using biofeedback, where they have been demonstrated through various usages ranging from educational support (Kato, 2010; Conati and Chabbal, 2003) to stress/relaxation/concentration training (Wang et al., 2011; Froment et al., 2009; Sharry et al., 2003). The motivation for serious games presented in this thesis arose from the discrepancy in research on serious games using biofeedback for improving emotion–regulation and decision–making performance, in contrast to the healthcare field.

2.11.3 Serious Games in the Context of Financial Decision–Making

There has been progress regarding serious games in the context of financial decision–making. These games may simulate real–life financial situations players will eventually find themselves in. The electrics company Philips together with the Dutch bank Abn Amro developed a biofeedback device named 'Rationalizer' for retail investors (Djajadiningrat et al., 2009). The device allows users to see the intensity of their emotions reflected in the form of dynamic lighting patterns. In recent years, the advances in cognitive neuroscience have also gained increasing interest from information systems researchers (vom Brocke et al., 2012;

Dimoka and Banker, 2012; Dimoka et al., 2011; Riedl et al., 2009). As outlined by Dimoka et al. (2011), the nascent field of information systems relying on neuroscience (NeuroIs) is drawing upon the theories, methods, and tools offered by cognitive neuroscience. This also includes psychophysiological tools such as Ecg and Gsr measurements (Riedl et al., 2009). Research in the field of NeuroIs has the potential to provide long–overdue insight into the decision–making processes of users interacting with information technology, where the possibility to map the underlying neural emotional mechanism and activation of Ans exists. For instance, Riedl et al. (2011) showed that deciding whether to trust an avatar induces less intense neurobiological processes than deciding whether to trust an actual person. In a different study, Riedl et al. (2010) reported that there are gender differences concerning the activity of specific brain areas when deciding on the trustworthiness of eBay offers. Based on Hr measurements, Adam et al. (2013) showed that in 'electronic Dutch auctions' the impact of time pressure on final prices is mediated by arousal.

2.11.4 Serious Games in the Field of Human–Robot Interaction

Previous studies investigating the interaction between humans and robots are sparse. Robots have been used: in a serious game setting for treatment of autism and stress (Tapus, 2013; Wainer et al., 2013); in a game scenario, interacting with groups of children (Walters et al., 2005); in games and other natural social interaction with humans conveying emotions and robots providing feedback (Brooks et al., 2004; Breazeal, 2002).

Physiological signals are employed in various application, but their application to humans in collaboration with robots on serious games in the context of a social interaction setting detecting underlying emotional states are limited (Agrawal et al., 2008). The following studies successfully implemented physiological measurements of elicited arousal using social cues in the Hri–enhanced serious games. The studies report an increased performance and higher positive valence between robot collaborator conditions for the lower arousal condition. Peck et al. (2014) used robotic basketball to learn the preference of the children with the autism spectrum disorder, where anxiety levels (as affective states) were implicitly measured using physiology. Pereira et al. (2012) designed an artificial opponent able to socially interact (using social, emotional and cognitive aspects of interaction) with multiple players in a digital tabletop game of Risk. Hirth et al. (2011) created a humanoid robot which had sensors (i.e., cameras, microphones, and kinematic system) to infer the emotional state of the humans interacting with it. Kim and Kwon (2010) developed a game of twenty–questions as an interactive task with the robot system apprising emotions. Tapus (2013) created a serious game called ‘Stress Game’ to elicit stress and thus frustration in the players, using the robot which continuously monitored the performance level of the players to infer their state of frustration. Dehais et al. (2011) used the robot to assess the human affective state in response to robot gestures. Esau et al. (2008) designed a robot able to recognize human emotions through the use of facial expression and react adequately by playing a game.

2.11.5 Serious Games for Practicing Emotion–Regulation through Biofeedback on the Decision–Making Performance

In the Frozen Bubble game (Parnandi and Gutierrez-Osuna, 2017) the player is presented with an arena containing a spatial arrangement of colored bubbles, and the goal is to clear the arena. Placing a new color bubble next to two or more of the same color makes them disappear; otherwise, they pile up until the arena fills up. Biofeedback is presented through the difficulty of the serious game, while the difficulty is controlled through the initial random arrangements of bubbles from one screen to the next, while the decision time pressure is controlled through the ceiling of the arena which drops one notch every eight moves and reduces the play area over time. In the Sno–Cat (Larkin et al., 1992) the player navigates a motorized vehicle up the mountain evading randomly–appearing trees on the way. The biofeedback was presented through the continuously updated performance score and the background color. In both of these studies, participants played a serious game coupled to their HR biofeedback presented, and gave evidence that the combination of indirect (covert) biofeedback presentation and direct (overt) visual biofeedback presentation in the acquisition and transfer of deep–breathing skills increase performance on a decision–making task. In the serious game based on a modified version of the Left4Dead game (Bouchard et al., 2012), the player is set as a survivor of a zombie apocalypse, who has to reach a rallying point. The difficulty of the game was controlled by the strength of zombies, while biofeedback loop also included obstructing the field of view and pre–recorded heartbeats increasing in pace and volume. The authors gave evidence that the biofeedback based on HR in serious games improves the emotion–regulation skills measured on salivary cortisol, which in turn increases performance on the decision–making task.

The rest of the serious games presented in this section follow a similar method of increasing the difficulty and presenting visual biofeedback for a goal of practicing emotion–regulation to increase the decision–making performance. Chapter 11 goes into details on these games and investigations, while the general overview is presented here. In the Space Investors game (Hilborn et al., 2013), players navigate a spaceship from one planet to another in space, while trying to avoid getting hit by the asteroids. In the Basic Math game (Bontchev and Vassileva, 2016), players must choose the correct mathematical operation by moving the falling object, obtaining points by dropping the object into the correct container before it reaches the bottom of the screen. In the Pong game (Changchun et al., 2009), players defend their side of the screen with a paddle changing size and speed, against an incoming ball that also changes speed and size based on biofeedback. In the Emoshooter game (Kuikkaniemi et al., 2010) players have to aim at and shoot as many far–away enemies as possible, which attack the player in large formations based on biofeedback. Pacman Zen (Zafar et al., 2018) is similar to the original Pac–Man game where players collect the white dots strewn across the maze. The difference in this version is that every ghost directly chases Pac–Man instead of enacting a more complex chasing behavior. With each new level, the speed goes up, increasing the difficulty. In Dodging Stress (Zafar et al., 2018), players guide the ball from one end of the screen to the other, avoiding obstacles in the way. The goal is to reach the target on the other end of the screen. The game also features one continuous level where the game difficulty increases each time players reach the target.

2.11.6 Summary

As demonstrated, there exists a number of serious game applications in various fields. The articles report implementation description and results, but at present, a unified approach to serious game development does not exist. Nor are there any standardized methods to evaluate games, where serious games pose a particular challenge. Physiological measurements emerged as objective data metrics available in the field of evaluating game development, as well as an important tool for game analytics (Drachen et al., 2013). Although physiological sensors have been used in game research in the past to explore players' reaction to gameplay events, the industry still has not embraced these methods from academia (El-Nasr et al., 2013). This vagueness of development methods is a problem to the scientific game investigations since it leaves the description space of game features on loose foundations. The ongoing work in the creation of game taxonomies, what Zagal et al. (2007) refer to as game ontology feeds into the development and evaluation methods proposed in this thesis (Lindley and Sennersten, 2008; Björk and Holopainen, 2004; Aarseth et al., 2003; Caillois, 2001).

Chapter 3

Approach

3.1 Aim and Scope

This thesis aims to investigate the process of designing and developing serious games which use psychophysiological methods for supporting emotion–regulation on decision–making tasks. At the time of writing this manuscript, there exists no unified method to develop and evaluate serious games, let alone serious games using psychophysiological methods. By providing a framework for developing and evaluating such serious game applications, the processes and the applications can be scientifically validated. Moreover, combining fundamental theories with different psychophysiological methods can help to recognize and map emotions in a more general and straightforward way, providing a method of incorporating them inside of such serious game applications. Such coupling could investigate the effects of both arousal and valence on the decision–making performance, and the contribution of both Pns and Sns to this effect. Following this conjecture, the scope of this thesis is limited to serious games for supporting emotion–regulation through physiological biofeedback on decision–making performance, communicating individuals' affective information.

3.2 Research Questions

Serious games have been used as an investigation environment in this thesis, to explore the effects of emotions and emotion–regulation on decision–making performance. Serious games can provide a context for the on–line perception of emotional responses, which are reflected on the changes in the individuals' physiology through biofeedback (Yannakakis et al., 2016). Furthermore, serious games can have a positive impact on the development of many different skills which are impractical in the real world for reasons of time, safety, cost, or logistics (Ellis et al., 2006; Corti, 2006; Van Eck, 2006; Mitchell and Savill-Smith, 2004; Squire and Jenkins, 2003; Rieber, 1996). Such biofeedback serious games provide a shortcut to raising awareness of internal bodily states and emotions, which are one of the key factors that critically influence human decision–making (Adam and Gamer, 2011; Loewenstein, 2000). Emotions and decision–making biases are even beyond conscious awareness, and it is therefore highly interesting to regulate such emotional dispositions (Fenton-O'Creevy et al., 2011). Therefore, experience and awareness of emotions are necessary prerequisites for advantageous decision–making (Biais et al., 2005; Bechara and Damasio, 2005; Gross, 1998b; Bechara et al., 1997). It is important to highlight that emotion–regulation skill–acquisition is 'maximized' when individuals are provided with engaging context–dependent serious games (Dror, 2008). Moreover, emotion–regulation can only be actively practiced when individuals face an arousing environment (Burgoon and Bonito, 2002).

It has been found that individuals aware of their emotional states avoid being overpowered by their emotions, and need to develop skills to interpret and down–regulate those affective states of high arousal (Fenton-O'Creevy et al., 2012a; Peterson, 2007). Emotion–regulation using biofeedback has been extensively investigated, but it is only by the introduction of biofeedback where emotional states are presented on–line coupled to serious games with immediate feedback regarding decision–making tasks, that the role of emotion–regulation in making decisions could be put in the context (Yannakakis et al., 2016). Evidence shows that emotion–regulation skills can reduce aversion of negative emotions and contribute to advantageous decision–making by helping to mitigate emotion–related decision biases (Fenton-O'Creevy et al., 2011; Heilman et al., 2010; Gross, 2009). Individuals with low emotion–regulation skills are prone to make disadvantageous decisions with undesirable consequences (Sütterlin et al., 2011). Following up on these findings, this thesis investigates the awareness of emotions, their regulation and how it influences decision–making. Therefore, the following research question takes into consideration the decision–making performance in serious games, taking into account the practice of emotion–regulation skill through biofeedback. This research question has been explored in Chapters 5–7 and Chapter 11.

RQ I. *What are the effects of rewarding and practicing emotion–regulation measured on the decision–making performance in serious games through biofeedback?*

The affective information provided by different psychophysiological modalities is diverse, especially as each modality provides a slightly different window to the activation of the ANS. Because of this reason and the information overlap between different modalities, selection of the optimal ones for a specific problem is not a trivial issue. Physiological modalities provide important insights into the internal emotional states, which are not readily available through other observational methods. Previous investigations gave evidence that diverse emotions evaluated through HRV are correlated with the patterns of ANS activity (Demaree and Everhart, 2004; Levenson, 1992). There is an affiliation between SNS activity and physiological arousal, and it has been closely linked to emotions (Scheirer et al., 2002). Moreover, stressors are often associated with an increase in SNS and a decrease in PNS activity. Therefore, Appelhans and Luecken (2006) state that the measure of the continuous interplay between SNS and PNS influences yields information about ANS flexibility and thereby represents its capacity for regulated emotional response. Furthermore, it has been generally proposed that positive emotions are associated with PNS (Fredrickson, 2001; Porges, 1995), while negative emotions are associated with SNS activity (Fredrickson et al., 2000). Emerging emotions have their neural correlates which unfold in the hard–wiring of the human brain. No theory of emotions could be explained on the whole if it does not introduce the neuroscientific aspect of their emergence and unfolding. Following up on these findings, this thesis investigates the activation of both PNS and SNS, and their contribution to the effects of emotion–regulation on decision–making in serious games. Therefore, previous research has been extended to take into consideration the physiological affect, taking into account the valence and arousal in response to the decision–making task. In that regard, the following research question investigates the activation of the different pathways of ANS, measured through the physiological modalities (i.e., EEG, ECG, GSR, PD). This research question has been explored in Chapter 4, Chapter 8, Chapter 10 and Chapter 11.

RQ II. *Is it possible to map the activation of* ANS *during an emotionally–arousing decision–making task in serious games?*

Emotions can have a positive or negative influence on decision–making performance; therefore they may contain relevant information for individuals (Seo and Barrett, 2007; Shiv et al., 2005; Bechara and Damasio, 2005). There is extensive evidence that emotions are a crucial factor that critically influences human decision–making (Adam and Gamer, 2011; Loewenstein, 2000). Individuals under the influence of high arousal may be overwhelmed by their emotions in a demanding situation and lose control over their actions, which can significantly threaten rational cognition and decision–making (Kuhnen and Knutson, 2005). Physiological arousal, whether caused by a positive or negative experience, may be equally (and some argue even more) important than valence for decision–making performance (Reeves and Nass, 1996). Following up on these findings, this thesis investigates the effect of emerging emotions on decision–making performance in serious games. In that regard, the following research question takes into account the elicited physiological valence and arousal in response to the decision–making task in serious games. This research question has been investigated in Chapters 6–7 and Chapters 9–11.

RQ III. *What are the effects of emotions on decision–making performance in serious games?*

3.3 Methodology

The research approach applied in this thesis is based on both quantitative and qualitative methods. These methods include true– and quasi–random experiments, as well as evaluative research.

3.3.1 The Implementation of Serious Games

The Auction Game was used as a serious game in this study to investigate the cognitive abilities of the participants and the physiological arousal effect on emotion–regulation and decision–making (Jerčić et al., 2012). The Auction Game has emerged as an investigation platform since it presents a challenging decision–making task to the participants, interconnected with an ECG sensor for the on–line assessment of physiological arousal and the biofeedback presentation based on HR values. An advantage of such a tool was that it measured reliable physiological arousal in a stressful environment (Astor et al., 2013b; Jerčić et al., 2012; Peffer et al., 2010).

Figure 3.1: Screenshot from the Auction Game (Jerčić et al., 2012).

The participants were presented with a decision–making task to calculate a mean value from three given price estimations, to be able to reach a buy or sell decision at the correct price, see Figure 3.1. To make a decision, the participants had to click on the buy or sell button on the screen. Price estimations were directly linked to the physiological arousal level, such that they deviated from the correct price with higher variance the more aroused the participant was. Thus, lower physiological arousal would make the variance of price estimations closer to the correct price, so that a buy or sell decision could have been less challenging. The serious game was linear, which meant that there was always just one possible correct decision to be made in each trial, and it was a reasonable challenge for the participants. Effective emotion–regulation skills would benefit the participants to make the decision–making task less challenging, as higher physiological arousal also reduced the decision time, to promote a greater challenge. Moreover, the task became more challenging at subsequent trials as the decision period was further reduced, forcing a quick decision. The estimations were individually presented on screen for a certain amount of time, from one second at starting levels and shorter as levels progressed. Price–estimation clouds appear at random places on the screen to make the player attentive. The player's goal in the Auction Game was to reach the highest level possible. To advance from one level to the next, the players had to make profitable decisions and earn enough money to reach that level's profit goal. Moreover, they had to reach the next level in a limited period; otherwise, the game ends.

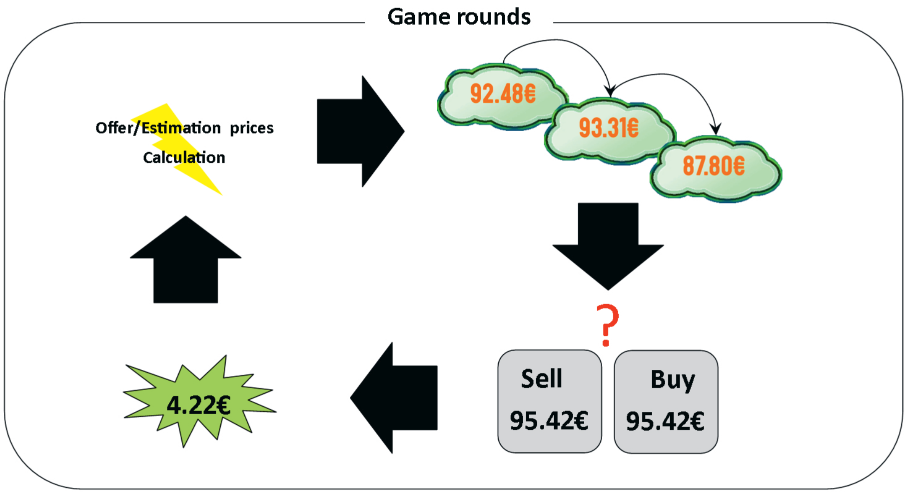

Figure 3.2: Example sketch of the game rounds (Jerčić et al., 2012).

The decision–making task was presented in the following steps, shown in Figure 3.2. Prior to the onset of price estimations, the participants had a one–second pause. Every price–estimation signal was displayed on the screen for one second, promoting fast–paced decision–making (Fenton-O'Creevy et al., 2011). The participants then had two seconds to reach a buy or sell decision, and shorter as the trials progressed or physiological arousal increased. After a decision, the feedback conveying the outcome of a decision was presented. The participants traded fictional 'goods,' and only the correct price was relevant for the participants. For each traded good the participants received three price estimations indicated on the screen. The mean value of these price estimations was the correct price for reaching a buy or sell decision for an offered price. Depending on a participants' decision, they could have earned or lost virtual money (on average 2.50 €) relating to a correct or incorrect decision. The decision had to be made within the allocated amount of time; otherwise, the participants were punished with a loss (-5 €). Depending on the implementation, a decision to pause for a turn and not make a trade was implemented. Pressing the 'no–trade' button is defined here as a correct decision since it always prevents against losing € 5.00 for not choosing one of the other options.

Level	Elements varying cloud estimations
1/7	Player's arousal is displayed but it has no effect on the game at all.
2/7	Player's arousal level affects the variance of estimates.
3/7	Estimation clouds move.
4/7	Estimation clouds become bigger and smaller in size.
5/7	Estimation clouds are same sized, but fake clouds with text start to appear.
6/7	Fake clouds with numbers start to appear. Distracting background images.
7/7	Estimation clouds become bigger and smaller in sizes again.

Table 3.1: Overview of the elements introduced for each game level (Jerčić et al., 2012).

The seven different levels of the game are summarized in Table 3.1. As the game progressed the difficulty changed, making it more challenging for the player to take a profitable decision. Moreover, from level to level, more and more distracting elements were included in the game environment. Among these are additional irrelevant clouds carrying false information, the reduced display time of the clouds, as well as auditory and visual distracting elements. Furthermore, the speed of the cloud movement was correlated to arousal at higher levels.

The Tower of Hanoi (ToH) game was used as the serious game for the studies in Chapters 9 and 10 to investigate the elicited human physiological affect in collaborative *human–human interaction* (Hhi) and Hri. Furthermore, it was used to bring such affect in relation to the performance on a collaborative serious game task (Jerčić et al., 2018b). Following the definition of serious games as games with a clear purpose other than entertainment (Susi and Johannesson, 2007), the ToH was considered as a serious game since it was used as a tool for the investigation of collaborative Hri. The game was easy for the robot to handle since an optimal solution to the game exists, and it was a reasonable challenge for most of the participants.

ToH has initially been a single player game. In collaborative gameplay, human–human or human–robot partners took turns to complete the game. According to the rules explained by Lucas (1893), the ToH is a mathematical game with disks of varying sizes that can slide onto any of the given pegs. The goal of the game is to reach from a starting configuration of the disks on the leftmost peg to the same configuration on the rightmost peg in a minimal number of moves (Lucas, 1893). In this study, the serious game started from a given configuration of the four disks, which was the same for all participants and referred to the beginning configuration in the game definition above. The individual trials consisted of moving any single disc to a next legal position, interchangeably between the participants and a collaborator until the final configuration of disks was reached on the opposite peg from the start. The participants always started first. Furthermore, similar to checkerboard games, implementing rules for the ToH was relatively simple (Xin and Sharlin, 2007). In games where social actions are not required (e.g., chess or checkers) and there exists only one human participant, autonomous agents that play optimally are possible (Pereira et al., 2012). At every move in the ToH game, just one possible optimal step existed to move a disk toward the final configuration. The participants always had an option to take this optimal step as their next move, while it was mandatory for the collaborators. Therefore, only the participants had an option to take the non–optimal step to move a disk in any other legal position, which would not necessarily lead toward the final configuration.

3.3.2 Experimental Setup

The true experiment constitutes a quantitative approach that tests the impact of a treatment/intervention on an outcome (Creswell, 2008). This approach requires that factors affecting the experiment can be controlled and can consequently be called a controlled experiment. True experiments use random assignment of study units (participants) to the experimental treatment or to a comparison group, to ensure that the study units do not affect the outcome instead of the treatment, in the manner of *randomized control trial* (Kampenes et al., 2009). True experiments are confirmatory data analysis methods focusing on the testing of hypotheses, as used in Chapter 7. Repeated measures design was also used where the same participant was tested under two or more experimental treatments or conditions (Robson, 2002). Such a quasi–experiment method was also used (e.g., the emotion–regulation study in Chapter 6) because it was not possible to randomly assign participants to groups for the analysis since they already had a preferred emotion–regulation strategy. Exploratory data analysis, or exploratory research, is used to investigate little–understood problems, visualize data, and develop questions and hypotheses used in confirmatory data analysis methods (Robson, 2002).

Robson (2002) states that a questionnaire is a common qualitative method giving a view on the aspects hard to quantitatively measure. Nevertheless, it provides quantitative data of some qualitative (subjective) inquiry. Securing a high degree of involvement by participants is aimed for by administering a self–completion questionnaire immediately after the treatment or the experiment (e.g., as used in Chapter 9).

Evaluation research explores the effect or effectiveness of some innovation, service, intervention, and approach, using a combination of quantitative and qualitative methods. Such evaluation does not necessarily qualify as research but characterizes as one when carried out using a principled, systematic approach (Robson, 2002). The most important approach to evaluation research is making comparisons between groups, together with other aspects, such as evaluating whether or not the intervention meets the goals and the needs of the users (e.g., as used in Chapter 6 and complemented in Chapter 7).

This chapter tries to summarize and give a general overview of the experimental methodology used throughout the studies in this thesis. Therefore, there will be an overlap to some extent between the text in this chapter and the subsequent chapters detailing each individual publication. The reason for the overlap is that the studies and the papers presented in this thesis are connected, and feed into one another. Studies regarding The Auction Game given in Chapter 6–8 (Jerčić et al., 2018a; Astor et al., 2013b; Jerčić et al., 2012), as well as, the studies regarding the ToH given in Chapter 9 (Jerčić et al., 2018b) and the still unpublished Chapter 10, will each share the common experimental methodology.

Overall, controlled experiments have been conducted in a laboratory setting. The lighting and temperature conditions were controlled in such a way that artificial fixture light was used throughout the experiments while the temperature was held constant at 23 °C ± 1 °C. The participants were seated in a chair with a fixed height and a predefined position. The height and position were constant during the experiment. The two experimenters were always present in the laboratory room to monitor the experiments, but they were completely hidden behind a screen.

Upon arrival, the participants were given general information about the experiment and the description explaining the serious games used. Their written consent was obtained, explaining physiological measurements and data confidentiality. When the participants agreed to take part in the studies, they signed an informed consent form.

A five–minute baseline recording of individuals' levels of arousal was acquired during the rest period before the serious games started, as explained in Sütterlin et al. (2011). Subsequent arousal and valence values inferred during the tasks were derived from the physiological measurements in relation to the initial baseline period, where changes in the measurement resulted with changes in computed arousal or valence.

Regarding the Auction Game, the chest band holding a physiological ECG sensor was attached. Participants were seated in front of and facing the Tobii T60 eye–tracker at a 50–60 cm distance from the screen. A constant conversion

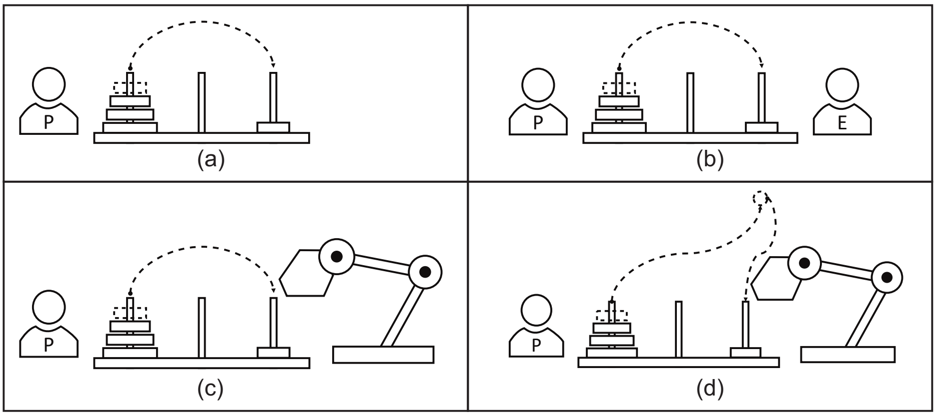

Figure 3.3: Experiment conditions (Jerčić et al., 2018b): (a) the *Solo* condition in which the participant is playing the game on his own; (b) playing with the *Human Collaborator* emulating the direct robot collaborator condition, with direct path at fixed speed; (c) the *Direct Robot Collaborator* condition, always moving in a similar fashion with a direct path at fixed speed; (d) *Non–Direct Robot Collaborator* condition which had one additional non–direct random point inserted in its path when performing the moves at varying speeds.

of game Euros earned into real cash was used according to participants' performance. Nevertheless, the participants received an equal base amount of Euros for participation in the study, which was paid directly after the game finished. The monetary reward was included in the ethical approval obtained through the xDelia project (FP7–ICT–231830). Participants were randomly seated in the laboratory, and the use of the sensors was explained during the registration. In order to reduce distractions, the participants were seated in their cubicles with noise–canceling headphones. After the instructions were given on the screen and read via headphones to the participants, they started playing a tutorial which walked them slowly through the game mechanics. Participants had to complete the training mode successfully before a five minute baseline period started and they could proceed with the game session.

Regarding the ToH game, the robotic experiments in this thesis were investigating the effect of physiological affect on the performance of the human and robot collaborator conditions in serious games. The main manipulations (see Figure 3.3) were: (a) the *Solo* condition where no collaborators were present and in which a participant was playing the game on his own; (b) the *Human Collaborator* condition emulating the direct robot collaborator condition, with a direct path at fixed speed; (c) the *Direct Robot Collaborator* control condition, where the participants played together with the robot which was always moving similarly with a direct path at fixed speed; (d) the *Non–Direct Robot Collaborator* condition which had one additional non–direct random point inserted in its path moving at varying speeds. The experiment setup was identical between the trials and participants, where participants played the turn–taking ToH game together with a robot or human collaborator. The goal of the game task was to move the disks from the starting to the final configuration. The collaborators were playing optimally on each move, following the algorithm. The human collaborator had been trained to interact the same way with every participant according to a well–rehearsed procedure.

The elicitation of physiological affect was achieved by the gestured motions and the speed of the collaborating robot. In particular, humans prefer that a robot moves at speed slower than that of a walking human (Butler and Agah, 2001; Karwowski and Rahimi, 1991). The gestured motions for the *Direct Robot Collaborator* were composed of a direct path at a fixed speed of 30 cm/sec (30% of robot speed) between the two endpoints of a current disc movement. A random path and speed between 5 cm/sec up to 70 cm/sec (5% to 70% of robot speed) were generated on–line for the *Non–Direct Robot Collaborator*. A non–direct path is generated using the two endpoints of a current disc movement in between which a random point in space above the disks was inserted, randomized on each game–move robot arm makes, which totals in three virtual positions the robot arm has to follow while making its move. The robot was passing through all the specified positions before having arrived at a final disc movement position.

The experimenters were present for safety reasons when using the robot arms and for monitoring that all data were recorded correctly. Surveillance of the robot arms and the participants was done using a live feed from a video camera. Also, the robot control software included an emergency stopping sequence that would trigger if defects in the program execution were detected, which was controlled at all times by one of the experimenters. Additional fail–safe software was introduced in the case of possible software failures. The participants were allowed for six minutes of rest in total between the trials. The four conditions (*Solo*, *Human Collaborator*, *Direct Robot Collaborator* and *Non–Direct Robot Collaborator*) were presented to each participant, as performed in the study from Burgoon and Bonito (2002). The participants repeated each of the four conditions three times one after the other (thus in all, a total of 12 ToH games were played per participant). Each experimental session took around 90 minutes to complete.

3.3.3 Experiment Procedure

Upon arrival of each participant, the following procedure was employed:

1. After entering the lab room, each participant was seated in a fixed chair at the table and faced the game task at a 50–60 cm distance.

2. The participants were given written information about the experiment and the description explaining the serious game. They were also given written information explaining that their data were stored confidentially. When the participants agreed to take part in the experiment, they signed an informed consent form.

3. Before starting the experiment session, the participants played a practice run of the serious game in order to acquaint them with the game task.

4. Before the experiment started, the participants filled in a demographics questionnaire. The questionnaire included familiarity with the game task, games in general and solving mathematical problems.

5. The physiological sensors measuring ECG, EMG, GSR, and EEG were attached, even though all of these physiological measurements were not used in the analyses. The participants were asked to relax in order to acquire a baseline recording of physiological data, during which the physiological sensors were calibrated.

6. In the experiment in Chapter 7, a true random experiment was used, and the participants were randomly assigned to the treatment groups. In other experiments, each participant performed all of the conditions of an experiment. The order of the conditions was counterbalanced between the participants in order to minimize the ordering effects and explore the difference in elicited physiological affect without the sequential effects. Each experimental condition was conducted as follows:

   (a) A participant played a serious game (i.e., The Auction Game or ToH).

   (b) After a trial was finished, a pause was allowed between the trial runs, where the questionnaire data regarding the gameplay experience were collected.

   (c) The operator instructed a participant what game task to perform next through the laptop placed next to the participant. A participant was allowed a rest of one minute before starting with the task trial. The operator controlled the laptop using a remote desktop.

7. Following the experiment, the physiological sensors were removed. The experimenter subsequently debriefed, rewarded, and thanked the participant.

3.3.4 Data Collection

The physiological signals were acquired using the Biosemi Active Two[1] physiological data acquisition system and its accompanying ActiView 9 software. ECG was measured at the chest using two 16–mm Ag/AgCl spot electrodes in a three–lead unipolar chest configuration. The right collarbone and the lowest rib on the left side were fitted with the two active electrodes, while the ground electrode was placed on the left earlobe. The signal was analyzed by software that detects R–spikes in ECG and calculates consecutive R–R intervals. ECG measures the rate and regularity of heartbeats by detecting the peaks of highly positive R–waves in the signal (Houghton and Gray, 2014). The beat–by–beat values were edited for outliers from artifacts or ectopic myocardial activity and linearly interpolated. GSR was measured using surface electrodes attached to the palmar surface of the middle phalanges from the middle finger and the index finger of the non–dominant hand (to reduce mechanical pressure susceptibility). This area, together with the palm and soles of the feet, is where eccrine sweat glands are most densely distributed, providing the strongest signal variations. Before the electrode application, the participants washed their hands with water and soap. The temperature and humidity were held constant across the sessions because of GSR susceptibility to the influences of temperature and humidity at the points where the sensors were attached. The sampling rate was fixed at 2048 Hz for all channels. Appropriate amplification and band–pass filtering were performed and subsequently down–sampled to 256 Hz upon data reduction.

3.3.5 Data Reduction and Analysis

The raw ECG signal was amplified, filtered using a high pass filter with a cutoff frequency of 0.1 Hz and a band–pass filter at 10–40 Hz, and 16–bit digitized, the signal was then smoothed by a 10 ms moving average window. The R–peak (heartbeat) detection algorithm analysis consisted of identification of R–waves applied to the ECG signal to obtain the interbeat intervals which were then reduced to HR, measured in beats–per–minute (bpm). The peaks were then detected in the resulting signal, and detection heuristic rules were applied to avoid missing R–peaks or detecting multiple peaks for a single heartbeat. These rules for obtaining the amplitude threshold (the difference between a peak and

[1] `http://www.biosemi.com`, accessed 07/02/2019 10:00

the corresponding inflection point) at which a peak is considered a heartbeat were: enforcing a minimum interval of 300 ms and maximum interval of 1500 ms between peaks; checking for both positive and negative slopes in a peak to ensure that baseline drift is not misclassified as a peak; and backtracking with reexamination/interpolation when a missing peak was detected. Such interbeat intervals were analyzed for ectopic beats (disturbance in the cardiac rhythm), arrhythmic events, missing data and noise effects outliers, which were linearly interpolated.

Regarding pupil data, off–line analysis of individual trials was carried out for the PD values in millimeters. The data were corrected for the short and long blink periods. A linear interpolation was then applied to the short blink periods. Baseline diameter was defined as the average diameter during the 60 s still image presentation. Baseline was first examined separately, after which it was used to normalize the pupil diameter data. Mean values for both the left and the right PD during the color still images exposure were calculated. The total minimum and maximum values for each pupil showed the range PD has among the set of the participants when exposed to the baseline stimuli. These results were used to form a normalized scale for subsequent PD analyses. PD maxima and minima were averaged, while the mean absolute values for each eye averaged across all participants for the baseline measurements for each pupil and each participant for the white, gray, and black screens. These values indicated where the center values were. The order of the presentations was counterbalanced between participants in order to minimize ordering effects.

Normalized PD values were grouped and averaged over the accompanying HR values recorded. The PD values were normalized for each participant using the BL_{min} and BL_{max} obtained from the baseline prestudy, using the normalizing Eq. (3.1).

$$PD_{norm} = \left(\frac{DP_t - DP_{min}}{DP_{max} - DP_{min}} \right) \times (BL_{max} - BL_{min}) + BL_{min} \tag{3.1}$$

GSR was measured in microsiemens (muS) and analyzed offline. Data reduction was performed using the Ledalab software for GSR (Benedek and Kaernbach, 2010). GSR includes short–term phasic responses to specific stimuli, and relatively stable, longer–term tonic levels (Andreassi, 2007). In continuous stimulus settings, the most common measures of this component are SCL and SCR, where their changes are thought to reflect general changes in autonomic arousal (Braithwaite and Watson, 2013). The authors state that SCR signal is suitable for assessing the intensity of single (phasic) emotions, but changes in the overall (tonic) level are relatively inert, thus valid for trials longer than two minutes. Changes in arousal within periods shorter than two minutes are not likely indicated using the SCL. This problem is particularly limiting in trials shorter than two minutes. When the SCL's temporal precision is insufficient, the rapidly reacting phasic changes NS–SCR seem to indicate a more likely focus. The number of NS–SCR during a given period is a prominent phasic–based indicator of arousal (Boucsein, 2012), such as a trial or a condition in these studies.

The raw GSR signal was down–sampled by a factor of 8 (from 2048Hz to 256Hz) to remove the high–frequency measurement noise and then smoothed by a 25 ms moving average window. Continuous decomposition analysis was performed. The phasic skin conductance detection algorithm used the following heuristics for considering a particular peak as a valid skin conductance response: the slope of the rise to the peak should exceed 0.05 muS/min; the amplitude should exceed 0.05 muS, and the rise–to–peak time should exceed 0.25 s. Because any function of SCR and its amplitude remains highly sensitive to the threshold that distinguishes it from noise, the standard threshold is set to a minimum amplitude of 0.05 muS (Boucsein, 2012; Dawson et al., 2007). Such noise may be a consequence of instruments' quality, the environment, or interpersonal differences. Once the phasic responses were identified, the rate of responses was determined. All the signal points that were not included in the response constituted the tonic part of the skin conductance signal. The slope of tonic activity was obtained using linear regression. Another feature derived from the tonic response was the mean tonic amplitude.

HRV is highly correlated with emotion (Lane et al., 2009; Appelhans and Luecken, 2006). Two measures of HRV are the SDNN and the LF/HF (Acharya et al., 2006). Wang and Huang (2014) state that SDNN and LF/HF are employed as two dimensions in the physiological valence/arousal model, where evidence reveals that SDNN is an efficient physiological indicator of valence (Kemp et al., 2011; Geisler et al., 2010). The total variance of HRV increases with the length of analyzed recording (Saul and Albrecht, 1987). Thus, it is not advisable to compare SDNN measures obtained from recordings of different durations. However, duration of recordings used to determine SDNN values (and similarly other HRV measures) should be standardized, to a minimum of 5 min recordings for short–term. Generally, SDNN levels for participants with positive affect are higher than for negative one (Geisler et al., 2010). For even shorter recordings, the main spectral components of LF/HF ratio are distinguished in a spectrum calculated from short–term recordings of 2 to 5 min (Malliani et al., 1991; Pagani et al., 1986; Akselrod et al., 1981; Hirsch and Bishop, 1981; Sayers, 1973).

Generally, the average change of HR is expected to range between 2–15 bpm (Bradley and Lang, 2000). The chosen interval threshold between peaks was well above the rate of change of Hr due to natural heart acceleration. Data were visually inspected for the artifacts which were subsequently corrected. R–R intervals were extracted. Time–domain features of the inter–beat interval, such as the mean and standard deviation, were computed from the detected R peaks. Inter–beat interval variability was explored by performing power spectral analysis on the data to localize the Sns and pns activities associated with different frequency bands. Ectopic beats, arrhythmic events, missing data and noise effects may alter the estimation of Hrv. Proper interpolation (or linear regression or similar algorithms) on preceding/successive beats on Hrv signals or its auto–correlation function may reduce this error (Kamath and Fallen, 1995). These steps were performed in Kubios software (Tarvainen et al., 2014), while variables of Hr, Hrv, Sdnn and Lf/Hf ratio were extracted in Hrv Toolkit[2].

3.4 Threats

Threats to validity can be divided into four main groups: external, internal, construction and conclusion, with each group containing several threats (Creswell, 2008; Wohlin et al., 2003).

External validity threats concern generalizability and ecological validity of the result. Since this thesis was evaluating a real–world variable in the designed lab environment, there was a possibility to miss on controlling all the aspects of that variable appearing in the real world. This would render the results of the experiment fault if inferred on the real–world setting (Wohlin et al., 2003). Tighter control of the experiment might help resolve this issue, but the participants are still left with the psychological pressure of being measured, where they might be influenced by the presence of the sensor equipment, leading to a distorted outcome of the experiment (Ouwerkerk et al., 2008). Authors further motivate the notion that this threat can be studied further in the future by introducing remote measuring methods.

Several possible internal validity threats can be ruled out by using randomized controlled trials in the experiment design (Robson, 2002). These include the selection of participants and the history of the experiment, but they have been dealt with by randomly selecting and testing all groups over the same period. Instrumentation threat has been ruled out by ensuring that the same instruments were used on all of the groups, even less so with repeated measurements design. On the other hand, this design introduces a maturation threat, as seen in the order effect where fatigue or practice can influence repeated measures at a later

[2] `http://physionet.org/tutorials/hrv-toolkit`, accessed 07/02/2019 10:00

point in the experiment. Especially interesting are the carry–over effects where there might be a specific effect on the second treatment, resulting from the particular nature of whatever is done first. Sufficient pre–experimental practice helps mitigate the issue, while counter–balancing the treatments helps mitigate order and carry–over effects. With questionnaires, there is a memory threat where participants might forget or alter their remembering of the treatment. This issue is reduced if the questionnaire is administered right after the treatment experiment has been carried out.

Construction validity threats are a result of inadequately defined and measured variables (Creswell, 2008; Wohlin et al., 2003). Both quantitative and qualitative variables measured in these publications are, within the domain, standardized and accepted measurements, with a long history of scientific use and validation.

Conclusion validity threats concern inaccurately drawn conclusions from the data, referred to as statistical conclusion validity (Creswell, 2008; Wohlin et al., 2003). Through the publications, standardized statistical tests are used to mitigate this issue.

3.5 Ethics

Apart from the validity threats listed in the previous section, there is an ethical aspect to consider as well. Ethical considerations are essential to any research involving people (Robson, 2002). Within such studies, there is usually the intention or possibility of behavior change. During the experiments participants were purposely faced with situations of stress and anxiety, so careful handling of the situation has to be practiced where participants were informed that they are free to stop and leave the experiment at any time. Participants have to be well informed about the nature of the experiment, up to the point of compromising the experiment. The data was securely stored, according to ethics and law regulation. These considerations are essential aspects of participants' right to privacy, dignity, and self–determination. Moreover, all experiments have to have approval from the ethical committee body revising the experiments in accordance with these considerations.

Specific ethical and legal issues must be satisfied before running an investigation with human participants. As suggested in Crown Copyright (2003), a risk assessment was carried out, followed with the considered mandatory activities (Walters et al., 2005). The participants in all of the experiments received complete information on the studies' goals, experimental conditions and gave their

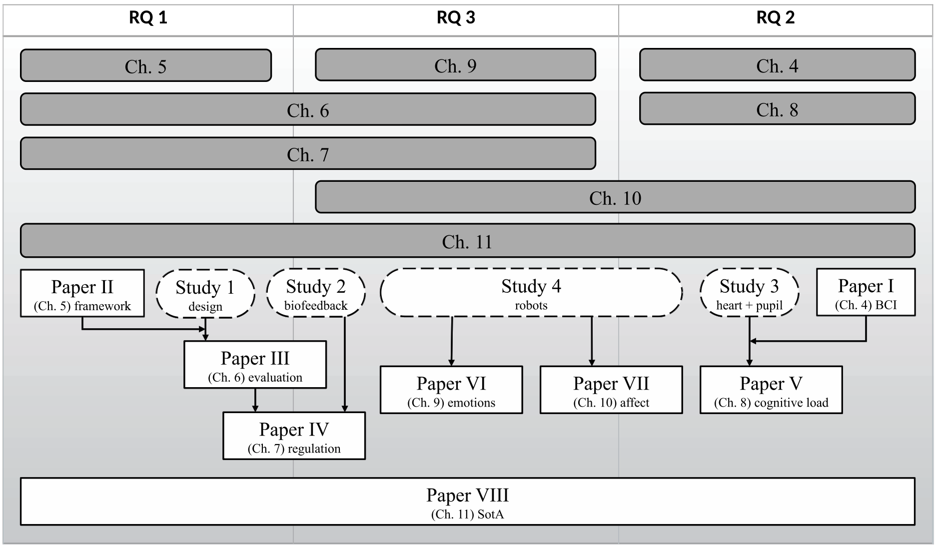

Figure 3.4: Overview of chapters (grey rectangles), studies (dotted line ovals) and papers (white rectangles) given in this thesis showing how each contributes towards answering the research questions (RQs) presented on the top. The arrows indicate how different studies and publications feed into the published research papers. Note that the order of RQs presented is intentionally shuffled to optimize the graph layout.

informed consent. The experiments were carried out by the Game Systems and Interaction Laboratory (Gsil) at Blekinge Institute of Technology, Karlskrona, Sweden. The experiments in this study were conducted under the advice of the local ethical committee, while the ethical approval was obtained through the xDelia project (FP7–ICT–231830). Moreover, the Ethical Review Board in Lund, Sweden, approved all experiments (reference number 2012/737) conducted in the PsyIntEC EU Project study (FP7–ICT–231143) through the European Clearing House for Open Robotics Development (Echord).

3.6 Contributions

The following chapter outlines the research contributions that have been published in the articles that constitute Chapter 4 through Chapter 11. Supported by the previous investigations surveyed in Chapter 11, this thesis brings together the biofeedback method and serious games to investigate the effects of practicing emotion–regulation. Furthermore, this thesis unites these concepts together for the novel application towards improving the performance on decision–making tasks. Figure 3.4 illustrates how each chapter, study, and paper contributes towards answering the research questions presented in Section 3.2.

Chapter 4 titled *The Future of Brain–Computer Interface for Games and Interaction Design* presents the potential of Bci paradigm as a non–invasive technology with minimal discomfort, where it has been successfully implemented and tested in various interactive applications and games up to date. Presented and discussed are potential application areas and uses of modern Bci technology on the example of an Epoc Eeg device. The discussion has been divided into two subgroups namely, game development and interaction design. Hence the future of these research areas in regards to such technology has been discussed. The contribution of this article is a review of interactive applications and games using Bci technology, in the attempt to present the current State–of–the–Art and infer the future direction. This chapter contributes to RQ II by presenting advantages and disadvantages in using Eeg compared to other peripheral psychophysiological modalities. For this publication, the thesis author was the main driver for the investigation, which comprised of reviewing the literature and writing the main body of the article. Some insights and additions to the text were done in cooperation with the co–author. This publication was presented at the Fun and Games 2010 conference workshop, Leuven, Belgium, 15–17 September 2010.

Chapter 5 titled *Evaluating Games Designed to Improve Financial Capability* presents the De toolkit capable of providing quantitative results for interactive applications and serious games, such as the ones presented in Chapter 4, set in the financial context. A multi–level approach to the evaluation of serious games for financial capability is presented in this chapter. The approach has been implemented as a toolkit in the context of xDelia, a collaborative project on Gbl with a focus on emotions in financial decision–making. The toolkit has been developed as part of a more extensive development and evaluation framework for the project. Four facets of financial capability games are targeted by the evaluation: game development, financial capability, behavior change, and learning with technology. The contribution of this article is an investigation into the process and metrics for developing and evaluating serious games set in the financial context. This chapter contributes to RQ I, in that it proposes practices, methods, and metrics for developing and evaluating serious games set in the decision–making context. For this publication, the thesis author was not the main driver, but he was involved in the investigation, which comprised of reviewing the literature and writing the article in cooperation with the other authors. This publication was presented at the 9th European Conference on eLearning (Ecel) 2010 conference, Porto, Portugal, 4–5 November 2010.

Chapter 6 titled *A Serious Game Using Physiological Interfaces for Emotion–Regulation Training in the Context of Financial Decision–Making* is in part a continuation of the work done in Chapter 5. In this article the De toolkit (Peffer et al., 2010) has been used to develop a serious game set in the financial context, aiming at helping decision makers improve their emotion–regulation. Research on financial decision–making shows that traders and investors with high emotion–regulation capabilities perform better in trading. 'Learning by doing' is a straightforward approach for participants to acquire the skill to regulate their emotions. However, this method becomes challenging without feedback. This problem particularly applies to the practice of emotion–regulation, because practitioners can get practically no feedback on their level of emotion–regulation. This research aims at providing an environment that can help decision makers to improve their emotion–regulation. The approach is based on a serious game with real–time biofeedback. The game is set in a financial context, and the decision scenario is directly linked to the individual biofeedback of the participants' heart rate data. More specifically, depending on their ability to regulate emotions, the decision scenario of the game continuously adjusts and thereby becomes more (or less) difficult. The participants wear an Ecg sensor that transfers the data via Bluetooth to the game. The game itself is evaluated at several levels. A small cohort of participants was typical for the experiments in this novel research field using demanding game artifacts designed. Therefore, careful consideration was taken to assure that the data was normally distributed for the analysis. The contribution of this article is twofold. First, it presents a serious game artifact using psychophysiological measurements for biofeedback, as discussed in Chapter 4. Second, it uses and evaluates the design and evaluation toolkit presented in Chapter 5, where results show that the developed serious game artifact met all the requirements of the environment supporting emotion–regulation. This study feeds into the subsequent work investigating further into the goals of this thesis. This chapter extends the answer to the RQ I and RQ III by evaluating practices, methods and metrics proposed in the design and evaluation toolkit; moreover, it evaluates the development process and the metrics for meeting design objectives. For this publication, the thesis author was the main driver for the investigation, which comprised of reviewing literature, developing the game artifact, setting up the experiment design, performing the experiment and analyzing the data. The development of the game artifact and writing the article was done in cooperation with the other authors. This publication was presented at the 20th European Conference on Information Systems (Ecis) 2012, Barcelona, Spain, 11–13 June 2012.

Chapter 7 titled *Integrating Biosignals into Information Systems: A Neuro*Is *Tool for Improving Emotion–Regulation*, is a continuation of work done in Chapter 6. In this article, the serious game presented in Chapter 6 and validated in

Chapter 5 using the De toolkit (Jerčić et al., 2012; Peffer et al., 2010), has been used to improve emotion–regulation in the financial decision context. Furthermore, this article evaluated the metrics of meeting design objectives proposed in the mentioned chapters. Traders and private investors are aware that emotional processes can have material consequences on their financial decision performance. However, typical learning approaches for de–biasing fail to overcome emotionally driven financial dispositions, mostly because of participants' limited capacity for self–monitoring (Biais et al., 2005). This research aims at improving decision makers' performance by (i) boosting their awareness of their emotional state and (ii) improving their skills for effective emotion–regulation. To that end, the serious game presented in previous chapters continuously displays the players' emotional states via biofeedback and adapts the difficulty of the decision environment to such emotional states. Subsequently, the developed serious game application was then evaluated in two laboratory experiments. Taken together, this study demonstrates how to improve financial decision–making by integrating tools from cognitive neuroscience in serious game applications. The contribution of this article is twofold. First, it evaluates the serious game artifact presented in Chapter 6, on meeting design objectives. Second, it presents an environment where financial decision makers can practice emotion–regulation. This paper validates the designed serious game and awareness of emotions in the moment of making decisions in neuroeconomic context. Therefore, it makes sense that two main drivers for this investigation were from the mentioned area. On another hand, the contributions of this paper were possible because of the serious game artifact developed; as well as, the experimental setup, analysis, and contribution to the text which were performed collaboratively with the thesis author. This chapter extends the answer to RQ I, by evaluating metrics for meeting design objectives; moreover, it contributes to the RQ III, in that it shows the effects on decision performance by rewarding and practicing emotion–regulation. For this publication, the thesis author was not the main driver, but he was involved in the investigation, which comprised of reviewing literature, developing the game artifact, setting up the experiment design, data analysis and writing the article, in cooperation with the other authors. This publication was published in the Journal of Management Information Systems (Jmis).

Chapter 8 titled *Modeling Cognitive Load and Physiological Arousal Through Pupil Diameter and Heart Rate*, is a continuation of work done in Chapter 4 which discussed an intent to use multimodal physiological interfaces to solve the current problems in cognitive science (Jercic, 2010). This article presents an approach to observing the Sns and Pns pathway activation and dominance over each other, using Pd and Hr as a measure of arousal. There is a need to develop applications tailored to individuals' needs and skills. Such applications would need to know when the individual users would perform optimally and when they would be overwhelmed, such that the task could be tailored individually for a specific user. This research explores the moment when demanding cognitive load overshadows the arousal effect on PD during a stressful task. To identify this exact moment, PD and ECG data were collected from participants playing a serious game. This dynamic environment was displaying biofeedback based on ECG measurements; furthermore, the PD data were analyzed in relation to the HR data during this stressful task. The peak of SNS activity was identified in the PD data; thus, this data coupling enabled observation of the SNS and PNS pathway activation and dominance over each other. The contribution of this article is the method capable of determining the activation and dominance of SNS and PNS pathway; moreover, the identification of the moment when overwhelming cognitive load overshadows the arousal effect. This chapter contributes to answering RQ II, in that it provides a method for mapping the activation of ANS during an emotionally arousing and cognitively demanding task based on the activation of SNS and PNS pathway measured using PD and HR. For this publication, the thesis author was the main driver for the investigation, which comprised of reviewing literature, developing the game artifact, analyzing the data and writing the main body of the article. Setting up the experiment design and some additions to the text were done in cooperation with the co–author. This publication was published in the Multimedia Tools and Applications journal. The publication was an extension of the paper titled *The Effect of Cognitive Load on Physiological Arousal in a Decision–Making Serious Game* published in the Proceedings of the 9th International Conference on Virtual Worlds and Games for Serious Applications (VS–Games 2017).

Chapter 9 titled *The Effect of Emotions and Social Behavior on Performance in a Collaborative Serious Game Between Humans and Autonomous Robots.* This article investigates perceived arousal–valence emotional and social cues in regards to robot collaborator modalities. It sets out to improve the development of serious games using socially–aware robot collaborators that can interact with their surroundings and other players on a social level. The notion presented is that robot collaborators might be more efficient or at least equally efficient as human collaborators for the sake of task performance. If one assumes arousal is more important than valence for decision performance purposes, developing a serious game using collaborator modalities that are more arousing, might be more suitable for use as a training platform. The contribution of this article is twofold. First, it identifies perceived emotions in the human–robot proximate collaboration in the same serious game, to understand what emotional and social cues underlie such an interaction from the human perspective. Second, it gives insights that could provide information for the development of more meaningful serious games that would use helpful and intelligent robots which act according to behavioral patterns that humans can understand and relate to. This chapter contributes to the RQ III, in that it identifies perceived emotions during a collaborative serious game and investigates how these affect decision performance. For this publication, the thesis author was the main driver in the investigation, which comprised of reviewing literature, analyzing the data and writing the main body of the article. Setting–up the experiment design and developing the game artifact was done in cooperation with the co–authors. This publication was published in the International Journal of Social Robotics.

Chapter 10 titled *Physiological Affect and Performance in a Collaborative Serious Game between Humans and an Autonomous Robot* is an extension to work done in Chapter 9 (Jerčić et al., 2018b). This article investigates physiological emotion correlates in regards to robot collaborator modalities, to improve the development of serious games using robot collaborators that can elicit a strong emotional reaction in human players. The evidence suggests that emerging emotions can influence rational decision–making, where they might be responsible for disrupting performance. This study attempts to understand what affective cues underlie human–robot proximate collaboration from the players' perspective in collaborative serious games, using psychophysiology. These insights could provide relevant information for the development of more meaningful serious games, which might increase decision performance. Nevertheless, this chapter suggests that, once optimal arousal has been reached, a further increase in performance may be achieved by regulating valence. Furthermore, if we consider that robots possess physical–virtual duality and have access to the game state information, possibly including psychophysiological measures, one can see why a choice of robot collaborator in a serious game would be a sound choice. The contribution of this article is twofold. First, it identifies players' physiological correlates of emotion on arousal–valence axes, in regards to the collaborating robot. Second, it gives insights into providing relevant information for the development of more meaningful serious games that would use robot collaborators which could elicit higher physiological arousal in players, improving motivation and performance of the human players. This chapter contributes to the RQ II, in that it provides a method for mapping the activation of ANS during a serious game based on the activation of SNS and PNS pathway measured using ECG and GSR. Furthermore, this chapter contributes to the RQ III, in that it identifies physiological correlates of emotions during a collaborative serious game and investigates how these affect decision performance. For this publication, the thesis author was the main driver in the investigation, which comprised of reviewing literature, analyzing the data and writing the main body of the article. Setting up the experiment design and developing the game artifact was done in cooperation with the co–authors. The publication was submitted to a peer–reviewed journal, and it was an extension of the paper published in the Lecture Notes in Computer Science, as the Proceedings from the International Conference on Entertainment Computing ICEC 2018.

Chapter 11 titled *Practicing Emotion–Regulation Through Biofeedback on the Decision–Making Performance in the Context of Serious Games: Systematic Review*, is a systematic, structured literature review to identify and investigate all the reported studies in *the effects of practicing emotion–regulation through biofeedback on the decision–making performance in the context of serious games* (BEDS). The evidence suggests that practice of emotion–regulation using biofeedback improves performance on a decision–making task in serious games through regulating levels of physiological arousal, where combined biofeedback presentation (i.e., direct visual and indirect gameplay) supports such emotion–regulation skill acquisition and transfer. The contribution of this article is twofold. First, it identifies the current state–of–the–art in the field of serious games using biofeedback for practicing emotion–regulation in the decision–making context. Second, it motivates the investigation topic of this thesis, situating it in the context of the current research interests. This chapter contributes to all of the research questions providing the literature survey and motivation for the findings presented through them. For this publication, the thesis author was the main driver in the investigation, which comprised of literature search investigation, reviewing literature, and writing the main body of the article. Quality and relevance coding of the papers was done in cooperation with the co–author. This publication has been published in the Entertainment Computing journal.

3.6.1 Summary of Contributions

It has been found that individuals aware of their emotional states avoid being overpowered by their emotions, and need to develop skills to interpret and down–regulate those affective states of high arousal. The steps to correct treatment have been validated in the previous reviews (Culbert, 2017), and they equal to: 1) awareness to objectively discriminate the mind–body differences between SNS arousal (stress) and PNS dominant states (relaxation response); 2) consistent regulation of a given psychological or physiological modality, towards a target direction (first in a controlled environment); 3) skill–transfer to apply this ability to regulate in the appropriate life–situation contexts, 'in the moment'; 4) change to restructure (CNS) and rebalance (baseline ANS) with regular daily practice to a long–term emotionally resilient pattern. Contributions from Chapter 6 validate the finding that the individuals were aware of their arousal through using biofeedback on a decision–making task in the serious game, which is a fist precursor step to employ emotion–regulation strategies to increase performance on a decision–making task. These contributions also take into consideration the screening of individuals for the targeted reappraisal emotion–regulation strategy, such that it becomes possible to make claims on its beneficial use for increasing performance in the main biofeedback serious game task. As traditional approaches fail in improving emotion–regulation skill–acquisition (Fenton-O'Creevy et al., 2012a), serious games provide a promising solution for common problems of traditional approaches, where direct on–line feedback on users' level of emotional–regulation and emotional states is available (Astor et al., 2013b). Contributions from Chapter 7 provide evidence to support serious games as a platform where the skill of emotion–regulation skill can be acquired and practiced.

Contributions from Chapters 6–10 provide the method and evidence for the validation of serious games, in regards to sufficient arousal elicited in the cognitively–demanding decision–making task. This allows the suppression/reappraisal emotion–regulation to be used to increase performance on the task, explaining the ambiguity of higher or lower arousal increasing performance. This thesis gives evidence that arousal is equally or more important than valence for decision–making performance, but once optimal arousal has been reached, a further increase in performance may be achieved by regulating valence. Furthermore, this thesis takes into consideration that the performance in the main biofeedback task was probably more difficult for the biofeedback group since they also had to learn how to regulate HR, which might have increased arousal on the task. Previous studies had coupled biofeedback target behavior to the performance, mainly since this was reflected in the game score, which also defines performance on the task. Without extensive training in emotion–regulation, this method of presenting the game score coupled to target arousal behavior through biofeedback presents another task to the users which increases difficulty. Therefore, it is hard to make inferences regarding the down–regulation of high levels of arousal increasing performance on decision–making tasks using biofeedback together with serious games. Following the propositions from previous investigations (Larkin et al., 1992), contributions from Chapter 7 provide an experimental setup using the sham and control condition groups to validate that using correct reappraisal emotion–regulation strategy leads to better performance on a decision–making task. Furthermore, such an experimental setup allows researchers to have comparable results between the control and treatment groups. Moreover, such an experimental setup allows making stronger claims that presenting biofeedback increases performance on the decision–making task, where biofeedback supports the practice of emotion–regulation in serious games.

The important point is that individuals integrate context–specific affective feelings into perception and their decision–making process. Carter et al. (2014) argues that the field of serious games is already recognized in research where decision–making questions are a burning issue. Therefore, contributions from Chapter 6 bring into focus how regulation of arousal is correlated with performance on a decision–making task in serious games, providing the context of fast–processing decision–making, where humans tend to rely on their emotions and follow an 'affect heuristic' without time for rational analysis. This has been achieved by introducing biofeedback presentation and performance measurement in the central decision–making serious game task. Neuroscience perspective might explain the skill–acquisition through using biofeedback in serious games, where performing goal–directed tasks (i.e., playing a serious game) leads to dopamines release in the striatum of the brain (Koepp et al., 1998). Therefore, serious games may facilitate improved awareness of arousal and effective skill–acquisition of emotion–regulation (Parnandi and Gutierrez-Osuna, 2017). Given the context dependency of skill–acquisition/transfer in the Beds context, contributions from Chapter 7 motivate the serious game context as a solution for the investigated problem, where the context was situated in the research domain.

Argued in Chapter 11, previous investigating in the Beds context gave evidence that the Ans might be responsible for the effect of different emotions on performance on a decision–making task. Nevertheless, these previous investigations do not explore further into the individual Ans pathways (i.e., Pns, Sns). Following the identified gap in the investigations of valence activation in distinct neural circuits, the contributions from Chapters 8 and 10 gave evidence for using different physiological measures to differentiate the Ans pathways (Pns, Sns). This proposition was followed in order to investigate how much these pathways participate in the arousal and valence affect. Moreover, the proposition was followed to also correlate these pathways with positive as well as negative affective states which are related to performance in decision–making. Therefore, this thesis evaluates the influence of affect on decision–making performance in serious games. As research in the Beds context is highly focused on the Ecg and inferring arousal using Hr, contributions from Chapters 8 and 10 provide the evidence motivating different biofeedback presentation modalities using other physiology measures, taking into account the player–centered design of such games (Culbert, 2017), where these modalities were made available through serious games in the studies presented in this thesis.

Contributions from Chapter 6 include extensive investigation in the evaluation and validation method for designing serious game artifacts for the research on the effects of practicing emotion–regulation through biofeedback on decision–making performance. Contributions from Chapter 5 include the (De) framework which proposes practices, methods, and metrics for the evaluation of serious games. The framework details the process of meeting the design objectives of serious game development. Therefore, this flexible framework supports serious game development and evaluation process.

3.7 Discussion

The techniques investigated in this thesis concerns the design and development of affective applications. More specifically, they explore serious games as interactive applications using psychophysiological methods to infer affective information from their users. Furthermore, these techniques investigate the effect of such applications on practicing emotion–regulation. Initially, methods and practices for developing serious games using psychophysiology were non–existent, and the main premise was to bring together knowledge from other domains of game development and affective computing. Using these methods for the first time in this context required extensive evaluation both of the methods and developed applications. As presented throughout the thesis, there were quite a few affective applications already available, but development or evaluation methods for those applications were not reported. Narrower search, focusing on the financial context, identified the lack of serious game applications using the mentioned methods. Following these notions, the DE framework has been proposed combining methods described above. It has been successfully utilized, both in the development and evaluation process of serious games, as well as, in meeting the design goals of such applications.

Regarding RQ I, psychophysiological methods have been a part of clinical and psychology research for a long time, but only using these methods on–line can one consider creating affective applications for the investigation in the domain of affective computing. Even though there was an abundance of information and methods from the previously mentioned fields which fed into the affective computing, it is only when using those methods on–line that new problems and questions arose alongside already known ones. As discussed throughout the thesis chapters, emotions are still as puzzling in science as they are in real life and the methods of detecting and classifying them are as unstable as the constructs they are trying to detect (Russell, 2003). More research is required as this area becomes acknowledged.

Emotion–regulation is a technique that helps improve performance on a specific task by controlling emerging emotions. The first step in practicing emotion–regulation is to gain awareness of those emerging emotions, so it makes sense to develop tools discussed in this thesis. The Auction Game and the ToH serious game were proposed and evaluated, with measurable outcomes presented. Such applications should be able to detect emotions to develop an environment that would present users with emerging emotions in order to raise awareness, as well as, to offer a possibility of practicing their regulation using the methods of biofeedback. Between the two emotion–regulation strategies, cognitive reappraisal was shown to be a more suitable one in a complex task of the Auction Game. The synchronous synergy between the users, affective applications, psychophysiological sensors, and methods is needed for such a goal to be successful. The results showed the benefits of practicing emotion–regulation in serious games, where decision–making performance was increased for the individuals who down–regulated high levels of arousal while having a positive–valenced emotional experience. Moreover, decision–making performance increased also for the individuals who received the necessary biofeedback information, where these individuals also had a significantly lower arousal level during the games session. The participants in the Auction Game who did not receive biofeedback information showed a discrepancy between the reported and the measured levels of arousal. Thus, the participants were able to receive information from the biofeedback and down–regulate their current level of arousal.

Regarding RQ II, psychophysiology offers a view on bodily reactions to certain emotions, but it is not until one dives deeper into the neural level of the ANS that one discovers the complicated and finely tuned mechanism of emotions, orchestrated by the brain. Performance on the task and the affect have two different ANS pathways competing over each other for dominance, PNS and SNS pathways. These implications have to be taken into account when developing affective applications for practicing emotion–regulation, sporting a cognitively demanding task. Only by understanding the underlying neural mechanism of emotions, one can truly begin to make general inferences on the observable effects of the human psychophysiology.

Regarding RQ III, the serious game applications developed in this thesis (i.e., the Auction Game, ToH) gave evidence that they are a useful tool to get participants aware of their arousal, as well as to help them in regulating their emotions to reach better decisions. The results from both the Auction Game and the ToH serious games showed the benefits of down–regulating high levels of arousal while having a positive–valenced emotional experience, where decision–making performance was increased for the individuals who had lower arousal. The uncertainty to whether higher or lower arousal is correlated with better performance on the decision–making task in serious games was tackled in this thesis by validating that the serious games elicited a sufficient level of arousal. The optimal level of stress and arousal for a given decision–making task depends on the context of the task and the individuals performing on it. Thus, it cannot just be said that down–regulation will increase performance. This thesis presents a player–centered design methodology and validates it on the decision–making tasks in serious games eliciting optimal levels of arousal for the cognitive abilities of the participants engaged in them. This method follows the notion that arousal is correlated with the performance score through a bell–shaped curve of Yerkes–Dodson–Law, which states that there exists an optimal arousal level for high performance on the task (Cohen, 2011). Without this step, claims cannot be made as to how arousal affects performance on the decision–making task in serious–games.

Proposed research questions feed into the overarching idea that a correlation between participants' emotion–regulation score and their average decision performance reveals a positive relationship between consciously employed emotion–regulation and decision performance in serious games. Furthermore, the results suggested that each emotion–regulation strategy has its benefits and shortcomings for the decision–making tasks. The Auction Game rewards effective reappraisal emotion–regulation with a high decision performance in the game and is thus suitable for helping users to practice emotion–regulation actively. Moreover, the results suggested that the interplay between PNS and SNS pathways could be mapped through underlying emotions activating these two pathways using psychophysiological methods in emotionally arousing serious games. Following up on these findings, through carefully balancing the activation of these two pathways, the results identified the optimal arousal level for increased performance of individuals engaged in decision–making tasks. Therefore, the moment where substantial cognitive load overshadows the physiological arousal effect was identified where reduction of the decision–making performance was observed through the cognitive demands of the task in the serious games. These findings may indicate the potential of serious games to provide game tasks tailored to the cognitive abilities of an individual interacting with them (Boutcher and Boutcher, 2006). Following the design and evaluation of the serious games presented in this thesis, informed serious games could be created which take into consideration an optimal level of arousal for a specific decision–making task. The investigations on ToH game also validated these findings in the collaborative serious game context, where robot collaborators were found to elicit diverse affect in their human partners, influencing performance on a decision–making task. The evidence presented suggests that arousal is equally or more important than valence for decision–making performance, but once optimal arousal has been achieved, a further increase in performance may be achieved by regulating valence. Furthermore, results showed that serious games designed in this thesis elicited high physiological arousal and positive valence, which makes them suitable as research platforms for the investigation how these emotions influence activation of PNS and SNS pathways, and influence performance on a decision–making task. Moreover, such serious games put participants at a highly aroused state where they need to practice emotion–regulation to succeed in the game. Such individuals who raised awareness of emotions and increased the skill of emotion–regulation were able to successfully regulate their arousal, which resulted with better performance, reaction times and attention scores on a decision–making task in the serious games presented in this thesis. As emotions are still a subtle and fleeing concept with substantial effects on decision–making, estimation and classification of those emotions are still challenges that are fueled by the physiological measurement variability between participants and longitudinal experimental sessions.

3.7.1 Development of the Serious Games

Jonassen et al. (2003) state that the complex approach to development and evaluation of technology–enhanced serious game should link goals acquired from empirical studies and evaluate the tools' potential to meet those goals. Therefore, such an approach should support interdisciplinary team learning and a shared understanding of research objectives, strategies, and activities underlying the design and development of serious games. A need to create such an approach in an iterative design and development fashion gave rise to the De framework which proposes practices, methods, and metrics for the evaluation of serious games. Chapter 5 details its development process and meeting design objectives. This flexible framework supports serious game development and evaluation process. At the time of writing this manuscript, there exists no unified framework to develop and evaluate serious games, even though heuristic evaluation and design frameworks have emerged (Shoukry et al., 2015; Omar and Jaafar, 2010).

Practices, methods, and metrics in the context of the De framework were proposed to design, develop and evaluate serious game applications. Knowledge of fundamental psychophysiological theories together with its diverse methods allows incorporating inferred emotional states inside of serious game applications. The development of such serious games has to match the objectives set during its design, which have to be evaluated for rewarding and practicing emotion–regulation. Developing serious games is a complex process involving collaborative cooperation between people from various fields. The process has to produce serious game artifacts that meet the design requirements. In case they do not, further development iterations are needed to get closer to meeting those requirements.

The De framework ensures that the comprehensive, ongoing evaluation is built into all facets of the project and that evaluation results find a way back into the ongoing development activities. This framework organizes the overall development and evaluation activities of the project and reports its findings back to the project. Furthermore, it encompasses functionality, heuristic and playtesting evaluation. Functionality evaluation during development consists of three categories: validation, verification, and future support. The heuristic evaluation aims at qualitatively identifying design errors and suggest improvements (Isbister and Schaffer, 2008). Playtesting evaluation evaluates the tool with the target audience. The De framework fulfills the role of a guide for designing effective project interventions by focusing on the problem of evaluation and design. Furthermore, it acts as a means by which to reflect on implementation processes and outcomes, involving the stakeholders as reflective evaluators and feeding the findings back into the project throughout the development. The evaluator uses this tool to rate the presence or absence of the different behavior–change strategies within five different levels of player interaction with the serious game. Those levels of interaction are information, skills, diagnosis, feedback, and gameplay.

The De framework differentiates between the two levels of design and evaluation, macro–level and micro–level. Macro–level De process thus organizes the overall design and evaluation activities of the project and reports its findings back to the project. For the more detailed specification of individual design and evaluation processes, the De framework provides a micro–level design and evaluation template. For a given project intervention, micro–level De process describes the different design and evaluation activities, together with any resources needed for their implementation. Specific parts of a micro–level De process may be sufficiently well–structured to allow for a stable representation through conceptual schemata or procedures and can be implemented as paper–based or software tools.

The Auction game, as a serious game developed in the frame of this thesis, was designed by an international consortium with scientific backgrounds in information systems, computer science, finance, and psychology. Since the xDelia project consisted of numerous application interventions similar to the Auction game, the collection of micro–level De processes has been a network of interrelated design and evaluation activities. In the context of the Auction Game, a set of specific requirements based on the existing research were defined which served as guidelines for the design process. Later these requirements were used as criteria for the evaluation of the design artifact. These criteria were: i) the design artifact has to be embedded in the context of financial decision–making and provide an engaging environment which can elicit arousal; ii) the design artifact has to provide an environment in which effective emotion–regulation is rewarded; iii) the design artifact has to provide the user with biofeedback in an unobtrusive and meaningful way. In the current version of the tool, evaluators assess the games regarding the 20 sub–rubrics and along the two or three levels of achievements, as proposed by Clough et al. (2010). A weighted score is calculated for each of the rubrics, where the weights vary as a function of the level of achievements. In the evaluation context of the Auction Game, development was followed by a heuristic evaluation aiming at qualitatively identifying design errors and suggest improvements (Isbister and Schaffer, 2008). A group consisting of four expert evaluators with years of experience in the game industry reviewed the prototype using heuristics divided into a set of categories, looking for potential usability and gameplay problems in the prototype. Heuristic evaluation pinpointed several important design issues and reinforced the robustness of the game. After which the playtesting evaluation study was administered. In the context of the Auction Game, the study confirmed that the game fulfills the average game usability, while it identifies an issue of how to optimally present the arousal information to the cognitively engaged player during fast–paced decisions. This issue was tackled through the identification of 'areas of interest' during the game sessions, and presenting the mentioned arousal information at these areas.

In the micro–level context of the Auction Game, the validation was done in collaboration between the partners from the xDelia project at several face–to–face meetings where the product was demonstrated. By letting the product owner test it before it was finalized, much of the somewhat abstract discussions became concrete as feedback was communicated. The verification was done by the team throughout the development process. By working in an incremental development style, the product was tested continuously for errors when additions were made, thus minimizing latent errors. Future support was achieved by designing the game around a modular and dynamic architecture, making it easy to adjust the game for future studies. Moreover, a separate component was used for measurement of arousal, allowing for other sensors to be used within the same game.

3.8 Limitations

The limitations of this thesis were several. The presented Auction Game was designed to reward down-regulation of arousal. While it was shown that in many scenarios arousal can impair decision–making (Peterson, 2007; Ku et al., 2005; Kuhnen and Knutson, 2005), there are also scenarios where it might be beneficial to up-regulate arousal to an optimal level and maintain that level (cf., Flow Theory (Nacke and Lindley, 2008) and Yerkes–Dodson Law (Hebb, 1955)). In this respect, it is important to note that this thesis does not pursue a normative approach in the sense that low arousal is always preferable. The serious game instead aims at improving the players' capacity to reduce high levels of arousal when awareness of it arises. The investigations in this thesis identified the optimal level of arousal where decision–making performance is highest. Therefore, future research should adjust the primary goal to target an optimal level of arousal in serious games in a more player-centered design. The serious game contexts presented in these investigations might have limited the generalizability towards any decision–making task context. Future studies would need to investigate further into the other task contexts.

The connection of game difficulty to arousal which also recursively affects the display of biofeedback is known as the 'biofeedback loop.' This loop makes it harder to draw inferences concerning whether immediate biofeedback also generally increases decision performance. In the Auction Game studies, alongside the evaluative approach that this context supported emotion–regulation practices, biofeedback was found to be meaningful in the sense that it helped players to perform better in the game. Future studies should investigate how biofeedback can be valuable decision support for the players and contribute to the long–term skill development of emotion–regulation in serious games.

The participants were entirely made of the college undergraduates and graduates. As serious games will be integrated into various contexts, future research should deepen the understanding of different populations interacting with the robot collaborators and serious games. Furthermore, there might have been fatigue effects present in this experiment, but it was assumed that they were minimal. This statement was supported by the results showing an increased performance for each participant during the later trials of the game tasks. The experiments lasted for about 90 minutes, with the time on the task around 20–30 minutes, which was comparable to other experiments reported.

3.9 Conclusion

RQ I is concerned with the observable effects on the task performance by practicing emotion–regulation with the help of serious games as supportive environments. The emotion–regulation strategy of cognitive reappraisal was investigated by using a serious game. The participants who used the mentioned strategy were more efficient regarding decision–making performance, over the others which have not. The effect of practicing emotion–regulation on a serious game has been investigated by presenting a stressful task, where groups with and without a supportive environment were measured in term of decision–making performance. The results of these experiments show the benefits of an emotion–regulation training environment on decision–making performance.

RQ II concerned the activation of two different pathways of the Ans on an emotionally arousing task. The underlying mechanism of emotions has been explored with the focus on Sns and Pns pathway activation. Using an emotionally arousing task, physiological measurements of Pd and Hr were evaluated to map this activation. Furthermore, the underlying mechanism of emotions has been investigated by mapping the interplay of Sns and Pns measured through Ecg and Gsr physiological modalities. The serious games used were found to be positively valenced and highly arousing. The results map the activation of the Sns and Pns pathways and identify the optimal arousal level where arousal effect on the physiology is shadowed by the cognitive demands of the task. This activation may be a useful indicator of the cognitive abilities of the participants in the context of a challenging decision–making task. These findings could provide a step towards understanding the activation underlying the Cns mechanisms, and their contribution to cognitive activities. Moreover, it may be beneficial to provide feedback for the participants to shape awareness of their cognitive abilities. Usage of physiological signals (biofeedback) in serious games could make possible the tailoring of serious games to the cognitive abilities of an individual participant, as previously suggested by Chiou and Wong (2008) and Charles et al. (2005).

RQ III is concerned with the measurable effect of perceived and inferred emotions on decision–making performance. This investigation was performed using a collaborative serious game with human and robot collaborators. While the effects of physiological emotions have been investigated and evaluated in the valence/arousal domain, the perceived emotions in the same context were investigated using questionnaires. The results from these experiments offer a possibility of designing serious games and robot collaborators, aware of the emotional state of players. The evidence from both serious games presented (i.e., the Auction Game and ToH game) showed that they elicited a sufficient level of arousal where the individuals with lower arousal performed better on a decision–making task. The evidence further showed that an additional increase in performance was observed for the individuals with more positive valence.

The aim of this thesis, guided by the research questions presented in Section 3.2, concerned the design and development of affective applications as serious games using psychophysiology methods to detect players' emotional states. The presented approach shows that it is possible to design and develop such applications that provide a supportive environment where decision makers can practice emotion–regulation and subsequently improve their performance on the decision–making tasks, which was chosen as a target focus for this thesis. Carter et al. (2014) argues that decision–making questions are a burning issue in serious games. If one considers that emerging emotions have a significant impact on decision–making, this motivates the notion that in the decision context, decision makers would benefit from regulating such emerging emotions. Furthermore, one can conclude that a supportive environment where emotion–regulation can be practiced is a beneficial approach for such individuals, where physiological methods provide the necessary affective input. In conclusion, by careful usage of evaluated design and development methods, affective applications for training emotion–regulation can be developed, with successful results improving decision–making performance in its users. To that extent, the De framework was proposed with its practices, methods, and metrics for the evaluation of serious games, supporting serious game development and evaluation process.

The contributions of the papers presented in this thesis are an extension to the previous investigations in serious games (Parnandi and Gutierrez-Osuna, 2017; Kim et al., 2013; Bouchard et al., 2012). These contributions include validation of sufficient arousal elicited in the decision–making task which allows the reappraisal emotion–regulation strategy to be used to increase performance on the task. As well as, validation that the serious games used in this study were cognitively demanding, suggesting a method of using physiological measures to differentiate the Ans pathways (i.e., Pns, Sns) in order to see how much they participate in the physiological arousal. This thesis provides evaluation and validation methods for designing serious game applications with biofeedback supporting emotion–regulation skill to be practiced on the decision–making tasks. It presents serious game artifacts which introduce biofeedback presentation and performance measurement in the decision–making context, where the coupling of biofeedback presentation and game difficulty has been automated. It also presents a method using serious games as tools for the scientific investigation using the context close to the researched problem, where sham biofeedback and no–treatment control condition groups have been used to validate and compare results between those groups. In the biofeedback experiments, sham (placebo) control condition is compared against the no–treatment control group, where the treatment condition group must be shown to be statistically superior to both sham and no–treatment conditions. Employing this method allowed for the investigation of the influence of emotions on decision–making by using biofeedback in the serious game task. Serious games were found to be a beneficial tool providing a correct context for the decision–making task, as this was shown to have a positive effect on the emotion–regulation skill acquisition.

Through the evidence presented, this thesis claims that practice of emotion–regulation using biofeedback improves performance on the decision–making tasks in serious games through regulating levels of physiological arousal, where combined biofeedback presentation (i.e., direct visual and indirect gameplay) supports such emotion–regulation skill acquisition. The serious games designed in this thesis have been evaluated and validated through the proposed methods, where evidence shows that these serious games are an effective platform for practicing emotion–regulation skill–acquisition and raising awareness of emerging emotions. It was demonstrated that the serious game applications developed through this thesis reward the participants achieving the desired arousal level, while at the same time presenting a hard challenge and penalty to the participants in undesirable high arousal states. Taking these findings into account, the serious games designed in this thesis allowed for the training of cognitive reappraisal strategy of emotion–regulation. The Auction Game provides the users with live biofeedback in a meaningful way which is, by design, based on unobtrusive physiological measurements. It has been shown that physiological arousal is equally or more important than valence for the decision–making performance, but once optimal arousal has been reached, a further increase in performance may be achieved by regulating valence. In conclusion, this thesis gave evidence that using evaluated design and development methods it is possible to design and develop serious games that provide a helpful environment where individuals could practice emotion–regulation through raising awareness of emotions, and subsequently improve their decision–making performance.

3.10 Future Work

This thesis presents a relevant question of emotion–regulation skill–transfer between contexts. Previous research gave evidence for the successful skill–transfer of emotion–regulation using biofeedback on the decision–making tasks in serious games. Nevertheless, it failed to provide knowledge regarding the context that is needed to make general claims (Parnandi and Gutierrez-Osuna, 2017; Bouchard et al., 2012; Larkin et al., 1992). Therefore, there is a need for an investigation regarding skill–transfer of emotion–regulation skills between contexts. Furthermore, since extensive emotion–regulation training was missing from the current research, the future directions should point out the apparent path of the long–term emotion–regulation training to further increase the decision performance. To that extent, future research should introduce participants who have already trained for a more extended period in emotion–regulation programs known to increase interoception (e.g., different forms of relaxation exercises, yoga, breathing patterns, or mindfulness courses), to evaluate the strength of emotion–regulation benefits to the decision–making performance.

The affective systems presented in this thesis follow the mainstream design logic which state that affective recognition systems today are highly specialized in using one psychophysiological modality, two at most for the most advanced ones. Even though these systems are used to modify the behavior of the applications using them, their primary objective is to model and map emotions in a research setting. This leads to high specialization, narrow approach and no generalization properties. This thesis presents a commonly reported problem of conflicting data incoming from different psychophysiological modalities. It is not clear if the problem emerges because of the cognitive side of affect models, or psychophysiological mapping and decisions algorithms. Future should witness a further separation of the stimulus and decision–making tasks one is trying to master, as was attempted in this thesis using collaborator conditions. This design enables investigations of affective states independent from the cognitive demands of the serious game task.

It would be interesting to develop an on–line affective recognition system with improved methods of mapping of Pns and Sns activation using psychophysiological methods and equipment. Such a system could communicate affective information based on neuroscientific principles into the application that the user is interacting with. The activation of mentioned pathways towards physiological arousal has been thoroughly investigated in the context of decision–making, while the future should witness a further investigation in the valence side of the emotional experience in this context, both on the cognitive side of the affect models and the side of decision algorithms.

Chapter 4

The Future of Brain-Computer Interface for Games and Interaction Design

Abstract

In this paper we discuss the potential application areas and uses of modern brain–computer interface technology such as the EPOC*. We divide the discussing into two subgroups namely, Game design and Interaction design and hence discuss the future of these research areas in regards to such technology.*

4.1 Introduction

The concept of brain–computer interface (Bci) is a relatively new field of research and it is not until recent years that equipment has made it possible to use the signals produced in the human brain as an actual vehicle for computer interaction. In this paper, we look at what has already been done in the field of Bci applications and then propose ideas on how to further extend that research and knowledge. Here, we focus mainly on the use of the Bci tools Epoc, developed by Emotiv.

The paper is divided into two main topics; Epoc for Game Design and Epoc for Interaction Design. We choose this separation because even though Interaction design might incorporate gaming and game design, it also represents a cluster of different research areas with very different purposes than those of game design.

4.2 Epoc for Game Design

Many games and applications today use biofeedback to help people to deal with emotional problems such as anxiety, phobia and post–traumatic stress. Since these types of disorders are all highly related to the brain, it is rather likely that brain–computer interface (Bci) technologies, such as the Epoc, could be very effective at detecting and communicating emotions to such an application.

Studies performed by Sharry et al. (2003) show that feedback regarding physiological arousal levels can help people to manage their feeling of anxiety as seen in Relax to Win (Sharry et al., 2003) where a game was created to help children who suffered from anxiety problems. By controlling stress levels, players relax more and thus perform better in the game. The level of the relaxation has been measured using Ecg and Gsr which have a lag time between the emotion and physical representation. We would like to introduce the Epoc to this type of game concept and upgrade the interaction between the player and the game. Since the Epoc is connected directly to the brain it does not have the lag times that some of the other sensory technologies have. With the Epoc we can detect early signs of tension and help change behavior before spiral arousal/tension has developed fully.

Using Bci as a tool to enhance interaction between the player and the game is an exciting concept. By feeding the game with Epoc sensory information regarding emotions and brain activity, we can modify the game play to drastically increase the game immersion, allowing people to play the game on a whole new level. Our suggestion is not to exclude other interaction devices or sensors, but to use the Bci equipment to supplement and enrich the interaction experience, much in the same manner as Froment et al. (2009). Here, participants play the game of Unreal Tournament while being connected to Ecg, Gsr and body temperature sensors. By feeding the game with this real time emotion information, the game is manipulated with effects such as screen blur and off–aim. The idea of the game is then to reward the player on successfully resolve the emotional issues. These types of concepts can enhance emotional self–control as the users are playing against themselves.

Introducing the wireless Epoc, we might be able to substituting multiple sensors, thus giving the player more freedom to interact with the game. Another aspect of the Epoc is that it can communicate many other things than just emotions and brain activity, such as facial expression and head orientation, all of which can make the interaction richer and more immersive. Also, in the game Brainball, Hjelm (2003) proves the potential usefulness of Bci in gaming by giving advantage to the most relaxed player.

4.3 Epoc for Interaction Design

Computers are oblivious to its users' emotional states. In order to create believable artificial intelligence, such as Hal in Kubrick's Space Odyssey, it has to recognize, interpret and communicate back emotions (Picard, 2000). By feeding emotional information back to the computer, we can make it interact with the user more naturally, recognizing the emotional state of the user and exhibiting sympathy and emotional intelligence which Picard (2000) states as a part of artificial intelligence. Using Epoc, we can directly communicate emotions, subtle facial expressions and brain activity directly to the computer making it aware of the complete emotional state of the person it is interacting with. The ability to communicate emotions is a trait of intelligent beings and as such by using Bci we make our computers intelligent. In this manner, we can make the computer pick up on emotional trends and act accordingly, learning and communicating with the user.

Using Bci as a tool to enrich interaction by communicating emotions makes for an interesting future. The potential benefits of communicating emotional states is clearly shown in FantasyA and SenToy (Höök et al., 2003) where a physical doll is used to express emotions back to the game by a set of body gestures. Players enjoyed this type of game interaction because they could communicate their feelings to the doll, which is the game avatar itself. The difficulty players had was the built–in indirect control mechanism between the doll and the avatar.

Using Epoc, we remove this indirect control variable and put more direct control of the avatar by using emotion and facial expression recognition. Players can then directly communicate their emotions to the game, making it less 'laggy' and frustrating. The stance of Pervasive Computing is to create context–aware applications that can adapt to information collected from the environment. Pervasive Gaming, however, deals with personalized aspects of how players are feeling at any given moment. As gaming moves more and more towards pervasiveness, we believe that the Epoc can communicate the player state as well as the environmental state to the application, thus enriches every application that uses this kind of pervasive interaction.

The Epoc is a rather non–invasive technology with no other discomfort to the user than that of wearing it on the head. As technology moves forward, so will this device and the possible feeling of discomfort will most likely be reduced. Brain–computer interface is an young and blossoming field of research with great potential and it is very exciting to be a part of such a young and blossoming field of research.

Chapter 5

Evaluating Games Designed to Improve Financial Capability

Abstract

A multi-level approach to evaluation of serious games for financial capability is presented in this poster. The approach has been implemented as a toolkit in the context of xDelia, a collaborative project on game-based learning with a focus on emotions in financial decision–making. This toolkit is as part of a larger evaluation framework developed during the course of the project. Four facets for financial capability games are targeted by the evaluation: game design, financial capability, behaviour change, and learning with technology. The development of this toolkit is work in progress. An evaluation exercise is planned with existing financial capability games, where we want to assess the toolkit and refine its design to make it more effective for evaluators to use.

5.1 Introduction

This poster illustrates the multi-level evaluation approach of xDelia[1], a research and technology development project funded by the European Commission under the 7th Framework Programme. xDelia is an interdisciplinary project that brings together experts from the fields of organisational behaviour, neuroeconomics, experimental psychology, sensor systems, experimental economics, cognitive sciences, game research, educational technologies, practice-based learning, financial capability, and investment banking. xDelia exploits new and emerging game and sensor technologies to investigate financial decision–making processes, including the role of emotions in people's decisions. Based on the insights gained from this research, the project will develop new, technology-enhanced approaches to financial training, with support for non-formal and informal learning in real world settings.

5.2 The Design and Evaluation Framework

The interdisciplinary character of xDelia, the multitude of interlinked empirical studies and game prototypes, and the international nature of the project's expert team requires an iterative approach to project implementation that is flexible enough to support the complex design and evaluation processes, and, importantly, to enable team learning and a shared understanding of research objectives, strategies, and activities. To ensure that comprehensive, ongoing evaluation is built into all facets of the project and that evaluation findings feed back into the ongoing development activities, xDelia has developed a Design and Evaluation (DE) framework tailored to the project's specific needs (Clough et al., 2009).

By focusing on the problem of evaluation and design, the DE framework fulfills a dual role. On the one hand, it acts as a guide for designing effective project interventions — workshops, studies, learning games, and so on — providing structure and support for good practice. On the other hand, it acts as a means by which to reflect on implementation processes and on outcomes, involving the stakeholders as reflective evaluators and feeding the findings back into the project on an ongoing basis.

The role of the DE framework is to structure and coordinate the different project interventions, making sure in particular that the information generated by the different design and evaluation activities finds its way back into the development process. We refer to this as macro-level design and evaluation. Macro-

[1] www.xdelia.org

level DE thus organises the overall design and evaluation activities of the project, and reports its findings back to the project. For the more detailed specification of individual design and evaluation processes, the DE framework provides a micro-level design and evaluation template. For a given project intervention, micro-level DE describes the different design and evaluation activities, together with any resources needed for their implementation. Since xDelia consists of numerous project interventions ranging from workshops to studies and prototypes, the collection of micro-level DE processes is in fact a network of interrelated design and evaluation activities, where each concrete DE process is structured according to the micro-level DE template shown in the centre of the poster.

Certain parts of a micro-level DE process may be sufficiently well-structured to allow for a stable representation through conceptual schemata or procedures, and can therefore be implemented as paper-based or software tools. The purpose of this article is to report on the first design iteration of a micro-level DE toolkit — a set of tools designated for a common use — for financial capability games.

5.3 A Design and Evaluation Toolkit for Financial Capability Games

The skills and knowledge, attitudes and behaviours that consumers need to enhance their financial security and wellbeing are generally referred to as financial capability (Kempson et al., 2006). The financial capability track of the xDelia project focuses on people's financial skills, attitudes and behaviours rather than on knowledge, and explores whether and to what extent serious games can be effective in this underexplored area. As a foundation for learning game design in financial capability, we have developed a DE toolkit that focuses specifically on the psychological determinants of financially capable behaviour. The goal is to streamline the DE process, making it more efficient and transparent, and to facilitate collaboration between domain experts and game designers. At the same time, we want to explore more systematically the possibility offered by game-based technologies to modify people's behavioural patterns and decision–making.

At present, there exists no unified method to design and evaluate learning games, let alone games for financial capability. Also, the idea of changing people's financial behaviour or target the psychological antecedents of maladaptive behaviour with learning interventions is quite new for policy makers and educators. Hence no well-documented initiatives exist to help us with the design of financial capability games. As a starting point, we chose a few topics that we thought our toolkit should address: financial capability, game design, learning with technology, and behaviour change. Each of these topics offers a useful and unique perspective on the game.

To speed up the game DE process, we have developed a set of spreadsheet-based evaluation tools that draw on existing models, frameworks, and taxonomies in financial capability, game design, learning, and behaviour change. The following sections describe the four different tools that make up the toolkit in more detail.

5.3.1 Financial Capability DE tool

A central aim of most financial education initiatives is to provide individuals with the necessary knowledge and skills to make informed financial decisions and appropriate choices. The financial capability DE tool evaluates the extent to which a game targets different financial knowledge and skills. To ensure that these are in fact relevant in the context of financial capability, we use the FSA/BSA Adult Financial Capability Framework (BSA and FSA, 2002) to develop the different measurement scales of the tool. The FSA/BSA framework, which proposes a set of key skills and competences that characterises a financially capable person, has guided the design of numerous financial education initiatives.

The current version of this spreadsheet-based tool groups the evaluation into three core sections — knowledge and understanding, skills and competences, responsibility — and nine sub-sections, which are further divides into 113 specific areas of knowledge, skills, and competences. Because evaluators may have problems to assess a game based on the rather vague short descriptions of the nine sub-sections, we have combined the over one hundred specific areas into a more manageable set of 34 categories. Currently, the tool allows evaluators to indicate which financial capability areas the game targets and how well it does this, and provide a weight for each of the evaluation areas.

5.3.2 Behaviour Change DE tool

While the financial capability DE tool primarily covers the more conventional financial skills, knowledge, and competences, what we are really concerned with in xDelia are games that in some ways target psychological and social factors that play a key role in motivating and influencing an individual's financial judgments and behaviour. Much of the psychological and behavioural research in economics has yet to find its way into developing better learning approaches for financial capability education. Behaviour change, including changes in antecedents of behaviour such as attitudes or perceived social norms, has been widely researched in health psychology. We turn to that literature to develop this De tool. In particular, we draw on standard behaviour models (Glanz et al., 2008) and on a behaviour change evaluation tool for physical activity web sites (Doshi et al., 2003).

In the current version of the tool, we have adopted the taxonomy of behaviour change strategies proposed by Doshi et al. (2003) — knowledge, cognitive strategies, behavioural strategies, and emotion-focused strategies — and added antecedent factors of behaviour found in standard models such as the health belief model and the theory of planned behaviour. Examples of such factors are self-efficacy, subjective norm, goal setting, feedback, and emotion–regulation. Although each of these factors can in principle be changed through learning or training, in practice this might be difficult to achieve or the effect on actual behaviour might be negligible. In general, because very little is known about effective ways to improve financial behaviour, the tool can only be indicative as to possible positive effects of a behaviour change strategy for an individual's financial capability.

The evaluator uses the tool to rate the presence or absence of the different behaviour change strategies within five different levels of increasing player interaction with the learning game. For instance, a game may provide the player with information about a behaviour change strategy, or it may help a player to find out about her maladaptive cognitions, or it may try to change a particular behaviour through game play. The levels of interaction are: information, skills, diagnosis, feedback, and gameplay. The tool calculates a score for the depth of player interaction, and we do not foresee at this point to rate the quality of individual strategies, since this is a complicated task that requires very specific expertise for each case.

5.3.3 Game Design DE tool

Badly designed games are unappealing, demotivating, and are unlikely to generate useful learning in individuals. At present, there exists no unified approach to game design — let alone to serious game design — that we could use as a basis for this tool. Currently there are no standardised methods to evaluate games, with serious games posing a particular challenge here. Here, we use Schell (2008) design lenses paradigm as a foundation for the game design De tool. This is a broad, practice-based approach to game design, incorporating a plethora of design aspect, such as essential experience, problem solving, competition, feedback, and so on. Each of these lenses ask a unique set of questions about the design, and we have modified them for the purpose of design evaluation.

Evaluating a learning game by asking focused questions about specific aspects of game design is highly intuitive and has a small learning curve for the evaluator, but still leaves room for expert judgment. Not all of the hundreds of questions of Schell's lenses are relevant for us at this point, and we have therefore reduced them to a more manageable size of 43 questions organised into eight categories: experience, game, elements, theme, iteration, player, mind, and mechanics, as advised in (Doshi et al., 2003).

This is the most complex of the four tools, with several auxiliary sheets to support the evaluation process. Each question, or aspect, is assessed on a five-point Likert scale and individual weights can be assigned to account for the importance of a particular aspect in a given evaluation exercise. Based on the rating and weight, scores are calculated for each of the five categories, and a total score gives an indication of the overall quality or merit of the game design.

5.3.4 Learning with Technology DE tool

Learning design, technology, content, and outcomes are important features of serious game design and evaluation. We are at a very early stage in terms of learning De tools, and at this point we have added only one tool for learning-specific design and evaluation in financial capability games: learning with technology. The tool draws directly on Jonassen et al. (2003), who proposed five attributes to characterise constructive learning with technology: active, constructive, intentional, authentic, and cooperative learning. So-called assessment rubrics are used to evaluate the extent to which both technologically mediated learning activities and the environment in which they take place promote meaningful learning in formal learning situations. Clough et al. (2009) adapted these rubrics to learning that takes place outside of the formal setting, and our goal is to modify them for learning with games. In the next design iteration we plan to add De tools for learning design and design evaluation, and for learning outcome evaluation.

In the current version of the tool, evaluators assess the games in terms of the 20 sub-rubrics and along the two or three levels of achievement proposed by Clough et al. (2009). A weighted score is calculated for each of the rubrics, where the weights vary as a function of the level of achievement. A first trial with these rubrics has shown that they will have to be modified in order to be operational and useful for serious game design and evaluation. Also, there are clear links between some of the sub-rubrics and the tools already present in other research areas, and these connections need to be determined and incorporated in later versions of the toolkit.

5.4 Conclusions

In this position paper we describe ongoing work on a design and evaluation toolkit for financial capability games as part of the xDelia project. The toolkit itself is developed in accordance with the design and evaluation process described in this article. We have now concluded the design part of the first iteration, where we developed four evaluation tools for financial capability, behaviour change, game design, and learning with technology respectively. The remaining work in this iteration is to evaluate the toolkit, which we will conduct using existing financial capability games. The outcome of this evaluation is recorded in the spreadsheet tools and provides the input to the second design iteration which will adapt and refine the tools. The second iteration will also provide us with an opportunity to extend the learning-related De tools, especially in terms of learning design and learning outcomes.

Chapter 6

A Serious Game Using Physiological Interfaces for Emotion Regulation Training in the Context of Financial Decision-Making

Abstract

Research on financial decision–making shows that traders and investors with high emotion–regulation capabilities perform better in trading. But how can the others learn to regulate their emotions? 'Learning by doing' sounds like a straightforward approach. But how can one perform 'learning by doing' when there is no feedback? This problem particularly applies to learning emotion–regulation, because learners can get practically no feedback on their level of emotion–regulation. Our research aims at providing a learning environment that can help decision makers to improve their emotion–regulation. The approach is based on a serious game with real–time biofeedback. The game is settled in a financial context and the decision scenario is directly linked to the individual biofeedback of the learner's heart rate data. More specifically, depending on the learner's ability to regulate emotions, the decision scenario of the game continuously adjusts and thereby be-

comes more (or less) difficult. The learner wears an electrocardiogram sensor that transfers the data via Bluetooth to the game. The game itself is evaluated at several levels.

6.1 Introduction

Serious games are (digital) games used for purposes other than mere entertainment (Susi and Johannesson, 2007). Corti (2006) points out an obvious advantage of serious games in allowing learners to experience situations that are impossible in the real world for reasons of safety, cost, time or logistics. Serious games can have positive impact on the development of a number of different skills (Ellis et al., 2006; Mitchell and Savill-Smith, 2004; Corti, 2006; Squire and Jenkins, 2003; Van Eck, 2006; Rieber, 1996) Having those different skills as defined learned outcomes, one can clearly see why serious games are considered as Game–based Learning (GBL) applications (Corti, 2006). To achieve the development of new knowledge and skills, game–based learning/serious games need to captivate and engage the end–users for a specific purpose (Corti, 2006). Corti further states that GBL has the potential of improving training activities and initiatives by virtue of, e.g., its engagement, motivation, role playing, and repeatability (failed strategies etc. can be modified and tried again); thus, lead to a more productive workforce.

Currently, there have been serious games created with the goal of teaching how to better manage financial decisions. These games may simulate real life financial situations players will eventually find themselves in, such as Massively Multiplayer Online Role Playing Game where the player has to make various financial decisions to gather enough money so they can retire and "win" (Jones and Chang, 2011) and "Darwin: Survival of the Fittest" where the player is thought options trading in the trading pit (Michael and Chen, 2005, p. 151).; or simulating business and stock market trading, such as a computer–based simulation business game where teams of players make various decisions regarding the product manufacture operations of a plant(s) and play the stock market trading company shares, where the team with the highest number of assets is declared the winner after the final period (Hartman and Galati, 2000). Also noteworthy to mention are the games of "Bankloan" and "Supra" where six players take the roles of representatives of three banks/companies seeking to trade loans and three supermarket buyers/sales–men each trading three products respectively (Abt, 2002, p. 101). These articles report students playing these games with enthusiasm as they were used in the curriculum; on the other hand, no measurable quantitative data has been reported on meeting GBL objectives.

Classical economic theories and models are usually based on the assumption of market actors being fully rational and favor utility maximization when confronted with economic decisions (Rasmusen, 2006). This way of considering economic decision–making has not only dominated economic literature for decades but has also shaped how humans perceive their economic decisions. In particular, professional investors and traders are considered to behave perfectly rational. However, the emerging field of *behavioral finance* gives broad evidence that not only financial amateurs, but also financial professional traders and investors suffer from strong decision–making biases (Shefrin and Statman, 1985). Especially periods of high stress and high market volatility can impair economic decision–making and hence trading performance (Lo et al., 2005).

There is broad evidence that emotions are one key factor that critically influences human decision–making (Loewenstein, 2000; Adam and Gamer, 2011). As will be shown in the next section, emotions do not always impair decision–making. They can also have a positive influence and facilitate decisions. Gross (1998b) argues that emotions act as response tendencies and subjects may follow these response tendencies or not. Recent research shows that, the ability of detecting or being aware of one's emotions and the skills to down–regulate levels of high emotional arousal improves human decision–making (Fenton-O'Creevy et al., 2011).

Following this conjecture, the international project xDelia (Xcellence in Decision–making through Enhanced Learning in Immersive Applications, www.xdelia.org) has developed a serious game that aims at improving the player's emotional awareness and training of his/her ability to down–regulate levels of high emotional arousal by the means of online information system displaying biofeedback based on *psychophysiological* measurements. As an advantage this system measures reliable emotional arousal in a stressful environment and is not biased by self–perception.

Based on the game, future experimental research can shed more light on the connection of training of emotion–regulation and decision–making. Moreover, a bank with high expertise in the private investors sector will test the game as a training tool for real traders and investors in day trading centers.

The remainder of the paper is structured as follows. In Section 2, we will first describe the theoretical background on emotions and decision–making upon which the development of the game is based. Section 3 describes how emotional arousal can be measured externally with the use of psychophysiological measurement technology. We then describe the design of our game approach — which we titled *Auction Game* — and present and discuss evaluation results in terms of game functionality and usability/playability.

6.2 Emotions, Emotion–Regulation and Decision–Making

Drawing from economic research there is broad evidence that economic decision–making can be biased to a considerable extent by levels of high emotionality and arousal (Loewenstein, 2000; Adam et al., 2011). In the context of economic decision–making, emotions are usually perceived as inappropriate interfering with the rational best decision and impairing the decision maker's ability to take "good" decisions. For instance, the disposition effect, i.e. the tendency for cashing in winning stocks quickly while holding on to losing stocks for too long, is often explained by subjects' emotional imbalance of how to cope with gains and losses (Shefrin and Statman, 1985). However, Seo and Barrett (2007) discovered in an empirical investigation with traders that emotions may also have positive effects on their stock trading performance. There is hence a bilateral effect of emotions: on the one hand they are bias–inducing and hence malicious to the decision maker, but on the other hand they also provide valuable knowledge in representing for example experiences one has gained in the past (Bechara and Damasio, 2005; Astor et al., 2011). Emotions can help evaluating situations instantly or processing informational overload, when one has to come to quick decisions.

As mentioned above, emotions also affect decision–making in professional settings, such as trading, which originally was believed to be a purely rational act. Several studies give evidence that professional traders are tremendously influences by their emotions. Fenton-O'Creevy et al. (2011) interviewed a set of traders in detail and reported that periods of losses were often accompanied by very risk–averse behavior and cautiousness. However, major gains often resulted in high confidence and headless behavior. These emotional states are often accompanied by high emotional arousal.

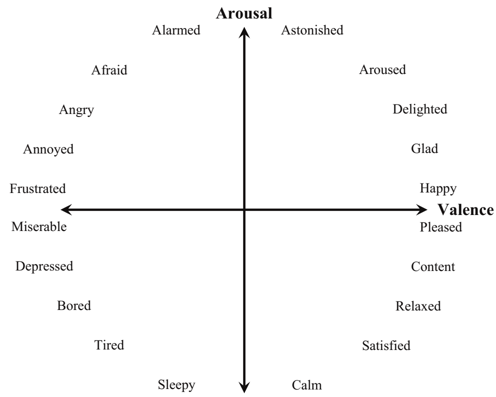

Figure 6.1: Emotions in the valence–arousal space.

Russell (1980) generally classified emotions by their independent components arousal and valence. Thereby, arousal represents the level of excitement whereas valence defines whether the current emotional state is positive or negative (visualized on Figure 6.1). Following this notion when measuring emotions, one is actually measuring a combination of valence and arousal. A reliable measure which shows different kinds of variation depending on the kind of emotional stimulus is the heart rate (Anttonen and Surakka, 2005; Vrana et al., 1986; Leng et al., 2007). Furthermore, since levels of high arousal can be accompanied by positive as well as negative emotions, arousal remains as the primary attribute of interest in our study. In the scope of our game approach we use heart rate as a proxy for emotional arousal, which will be described later. The continuous measurement of heart rate helps to improve the understanding of the emotional processing in economic decision–making.

Fenton-O'Creevy et al. (2011) further detected a strong link between traders' ability to regulate their emotions and their financial performance. The authors found that high performing traders have a better perception and awareness of their emotional state. Most interestingly, these traders are also more advanced in regulating their emotions. While less experienced traders usually try to avoid aversive emotions, the more experienced traders had actually learnt to cope with their emotions. Consequently, the more experienced traders were able to identify and discriminate their emotions in a more sophisticated way. Thus, there are interpersonal differences considering the experience, the awareness, and the ability to regulate emotions, which in turn inhibit or facilitate decision–making performance.

Emotion–regulation can be described by the process model of Gross (1998b) which is widely known and acknowledged in the field of emotion–regulation strategies. It relies on the assumption that emotions are generated in an emotion generative process. A broad distinction which the author draws is the one between antecedent–focused and response–focused emotion–regulation strategies. Antecedent emotion–regulation strategies apply while the emotion is still unfolding and has not reached its peak. An example for emotional *reappraisal* would be a shy student in a class. Now, emotional reappraisal could result therein that s/he considers the school class as a good opportunity to train raising his/her hand and answering questions. Hence s/he constructs a potentially emotion–eliciting situation in nonemotional terms. Response–focused emotion–regulation on the other hand tries to aim at altering and controlling the experiential, behavioral and physiological response to the fully established emotion. An example for such behavior could be the shy person in a school class which might try inhibiting ongoing emotion–expressive behavior and disguise them with e.g. insubordination.

People tend to use one of two main, broadly defined, strategies to deal with emotions emerging when facing difficult and stressful tasks (Wallace et al., 2009). These strategies are reappraisal and suppression. In line with the description in the previous paragraph, suppressors tend to constantly push down emotions, ignoring the fact that they exist and are continuously affecting them. On the other hand, reappraisers tend to positively re–evaluate situations. Both emotion–regulation strategies take up cognitive resources (Wallace et al., 2009). However, the authors also state that suppressing emotions generally takes up more cognitive resources in comparison to the reappraisal strategy when encountering undesired emotions. Hence, emotional suppression can eventually take up so much cognitive resources that it can reduce one's performance in decision tasks compared to the strategy of emotional reappraisal.

Gross and John (2003) designed a questionnaire — the Emotion–Regulation Questionnaire (ERQ) — in order to identify suppression/reappraisal strategy tendencies used by individuals. It makes specific statements with respect to the emotion regulatory process intended to be measured, such as "I control my emotions by changing the way I think about the situation I'm in."

This section has shown that emotions in the context of financial decisions can be both, bias–inducing, and performance boosting. Better emotion–regulation strategies result in better financial performance, whereby the awareness of the emotional state seems to be critical for appropriate evaluation of the decision situation. The developed serious game for our study thus aims at improving emotion–regulation, but also to improve the players' emotional awareness.

6.3 Psychophysiological Measurement of Emotions

In order to make players aware and give sufficient feedback on emotional arousal, it is crucial to apply a method to reliably detect emotional arousal. While subjective measures, such as self–evaluation, always also incorporate potential self–deception, we make use of psychophysiological correlates of emotions. Moreover, psychophysiological signals can be seen as an objective measure of the emotional state as it is hard to manipulate them intentionally. Iancovici et al. (2011) gives an up–to–date overview on (serious) games using biofeedback, where we can see various usage ranging from educational support (Conati and Chabbal, 2003; Kato, 2010) to stress/relaxation/concentration training (Froment et al., 2009; Sharry et al., 2003; Wang et al., 2011). There are a number of serious games in the field of healthcare and well–being using biofeedback (Kato, 2010), they are mostly concerned with hyperactivity disorder, autism, substance abuse (Wang et al., 2010) and more specific targets as pain, astma, bladder control, medical education on cancer. These games use various psychophysiology sensors to make the user aware of his/her medical state and provide a clear goal on how to improve it using feedback (Dunwell et al., 2010). This last point gives the motivation for the game presented in this paper; even more so if we consider that this overview describes the field of finance and financial decision–making as lagging behind in using serious games with biofeedback, in contrast to the healthcare field.

For computation of heart rate (Hr), we used the ekgMove sensor developed by Movisens which records electrocardiographic (Ecg) signals with high accuracy. In contrast to most other commercially available Ecg devices, the sensor is attached to the chest using a flexible belt with dry electrodes. Therefore, it is less obtrusive than other devices and offers a higher wearing comfort. The Ecg signal is transmitted via Bluetooth to the xAffect software environment (Schaaff et al., 2011). The software offers a modular framework which allows to process data from various input devices and to transfer the derived values via Tcp/Ip to other applications like the Auction Game. To get information about the current arousal level of a person, the heart rate is computed from the raw Ecg signal. An algorithm to derive the current arousal level from heart rate information is implemented in the xAffect framework. The arousal levels are computed in relation to a baseline period which is recorded before the game starts.

6.4 Game Description

6.4.1 Underlying Principle

The developed game serves two major goals:

- Improvement of *introspection*, the examination or observation of one's own mental and emotional processes, and self–monitoring of physiological arousal and hence personal emotional state.
- Improvement of skills in *emotion–regulation* by elements that reward good emotion–regulation and punish poor emotion–regulation strategies.

In order to achieve these goals, the game uses a *physiological interface* detecting online physiologically measured levels of arousal, as a basis for providing emotional feedback (biofeedback); furthermore, the game difficulty is connected to the measured level of arousal. The better the player is able to control and adapt his/her level of arousal, the easier the decision environment is.

The core motivation for the Auction Game is that there is a link between maladaptive financial behavior and poor emotion–regulation abilities. Therefore the Auction Game can be considered as an emotion–regulation training game in the context of financial decisions.

6.4.2 Game Concept and Gameplay

The narrative in the Auction Game is purposely simple, since it has to be easy to use the game for students, as well as for investors in day trading centers. The theme of the game is an abstract one depicting sky and clouds, as a supporting environment for down regulating levels of high emotional arousal.

The player is set in the position of a trader where s/he continuously can buy or sell goods, in each round one at a time. The game starts with the introductory screen where the player is presented with the instructions on how to play the game. Here it is also possible to go through the tutorial or just start the game, after which the player's baseline HR data is recorded. Every further arousal level measure in the game is calculated against this baseline. A previous tutorial explains the principle of the game and slowly guides the player. The tutorial should be played the first time the player gets in contact with the game, but it can be skipped if the player already knows the Auction Game.

Before the start of each round, an offer price and price estimations are calculated with respect to the level of arousal the player is currently at. A round

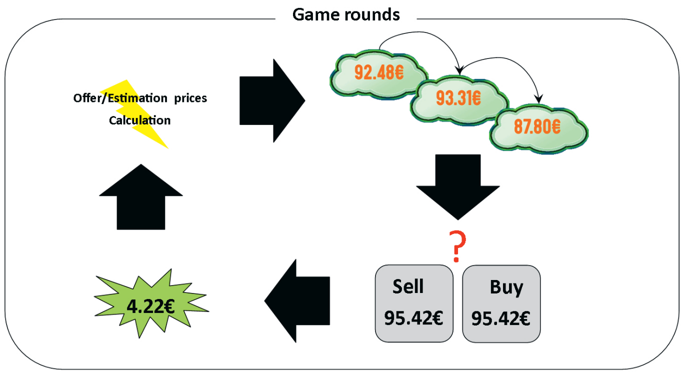

Figure 6.2: Example sketch of game rounds.

(see Figure 6.2) consists of three price estimations presented to a player sequentially, from the three trusted simulated consultants. The clouds are individually presented on screen for a certain amount of time, from one second at starting easy levels and shorter as levels progress. To make the player attentive, clouds appear at random places on the screen. They estimate the goods price in the next round; thus, by calculating the mean of the three prices in the clouds, the player knows the true price of the good/stock in the following round.

After the indication of three price estimations, the player gets the chance to buy or sell the good for the offered price. S/he has to make a decision in a certain amount of time, from two seconds at easy starting levels and shorter as levels progress. To make a decision, the player has to click on the buy or sell button; following this, an audio and video feedback is presented conveying the outcome of the decision.

Depending on his/her decision the player realizes a gain or a loss. In order to gain money the player has to take the right buy/sell decisions. Every profitable decision will reward the player with a certain amount of money, while a non–profitable one will reduce the player's earnings and take him further away from the current level's profit goal. Not taking a decision is the most expensive action, taking a large sum of profits (5 Euros) away from the player. Too many money losses will lead to the end of the game. This limit has been set to 10 Euros below the current level starting point. If the player is quick in calculating the three price estimations, s/he can easily reach a correct decision. After that, the total money earnings are updated and if the player has reached the current level's profit goal, a new round begins.

Consider the example in Figure 6.2. Assume that the price estimations are 92.48 €, 93.31 € and 87.80 € and that the offered price is 95.42 €. By calculating the mean the player realizes that the mean is close to 91 €. In fact, it is 91.20 €. Since the mean price of the estimations is lower than the offered price, the player should select the Sell button and realize a gain of 4.22 €. This task continues until the player has reached an upper bound to move to the next level or the player loses the game due to running out of time or due to bankruptcy.

The player's level of physiologically measured arousal affects the game difficulty. Before the start of each level, player is informed on how his/her level of arousal will influence the game difficulty. For example, the goal within a round could be to keep the level of emotional arousal as low as possible. As long as the player is able to keep arousal low the game will remain in the relatively easy mode. However, as soon as s/he becomes more aroused due to, for example, anger of an incorrect decision, the arousal bar will move up and the decision task will become more difficult. The level of difficulty increases by increasing the variance on the price estimation signals. While the price signals would normally (without arousal) be 92.48 €, 93.31 € and 87.80 € they could be then 68.22 €, 79.21 € and 126.17 € when the player is unable to keep his/her arousal down, making it more difficult to calculate (or estimate) the mean. The variance of the estimations will get larger the higher the emotional arousal is.

The player's goal in the Auction Game is to reach the highest level possible. To advance from one level to the next, the player has to make profitable decisions and earn enough money to reach that level's profit goal. Moreover, s/he has to reach the next level in a limited period of time otherwise the game ends; this time limit is currently set to 4 minutes. In every level the profit goal increases by 30 Euros. As soon as the player earns this amount, a button pops up and s/he can proceed to the next level.

Players who can achieve targeted the emotional arousal level will be rewarded with larger profit/lesser loss money values. On the other hand, undesired emotional arousal values will yield lesser profit and larger loss. While in the first levels the constraint for physiological arousal is to simply down–regulate arousal, in later levels the players have to aim at a specific level of arousal. As the game progresses, as well as arousal gets higher, the game difficulty will change, making it more difficult to take a profitable decision. Moreover, from level to level the tasks to regulate emotions will become more and more difficult as step by step new distracting elements are included. Among those are additional irrelevant clouds carrying false information, time, distracting images, auditory, and visual constraints (will be described later in detail).

Figure 6.3: Screenshot from the Auction Game.

The game is conceptualized such that it has no predetermined ending level, but after level 10 the game becomes extremely hard entering what is called "The Death Mode". Hence, the ending time of the single levels and the game can vary, depending on the player's skills. Optimally the game should run for approximately 25 minutes where players should earn around 200 Euros profit in the game. The player's skills to earn money in the game are related obviously to (1) his/her abilities to perform calculations under stress, and (2) to the player's skills to regulate his/her overall emotional arousal state. Independently from the player's calculation skills, good emotion–regulation will help to improve his/her individual game performance.

After the game has ended, the player is presented with the level s/he has reached. Better players earn a higher place on the high score list where they are given a chance to compare money earned result to their previous ones.

The objective of the Auction Game is to train players in performing emotion–regulation strategies. By showing the level of arousal the player can gain an awareness of his/her emotional state and the influence of emotions and emotion–regulation on decision–making. In other words, guide the player towards mindfulness of emotions. By displaying the player's emotional state as an indicator indicating arousal levels ranging from relaxed to highly aroused, a player has to regulate his/her arousal to minimize the deviation of the estimations and thereby have a better chance to accomplish a higher profit. Indicator levels are dependent on a player's level of arousal.

6.4.3 Game Mechanics

The Auction Game has been developed in a Unity 3D pro game engine which supports integration of third party APIs. As can be seen in the Figure 6.3, the Auction Game is played in a 2D environment where price estimations are presented inside of the colored cloud drawings. To depict a sense of progress through the game, every level has a different background picture of the sky. The player can see his/her individual arousal level indicated on the meter in the top right corner, as well as by the color of the clouds (green, yellow and red). The profit goal and total money earned are presented on the meter at the bottom right side of the screen. Decisions can be made by clicking on one of the Buy/Sell buttons presented at the bottom of the screen using a mouse.

6.4.4 Game Difficulty

The game must engage in play all types of players, ranging from experienced gamers to completely inexperienced players since the target group for a serious game may not necessarily be experienced game players. Thus, the game starts slowly introducing distracting elements step by step throughout the levels; moreover, the clouds appearance time (mean calculation) and decision time decreases.

In the Auction Game, different game elements are affected by the player's arousal level, which will make the game harder. In order to train emotion–regulation during the game, it is important that the game is sufficiently challenging in the aspect of emotional arousal control. The arousal affected elements described below are the different ways in which arousal influences gameplay, and are meant to make the game more difficult in different aspects. The further away the player's current level of arousal is from the target level, the bigger each of the effects will be.

Thus there are two dimensions of variety of difficulty in Auction Game respectively:

- **Game elements not affected by arousal**
  - As soon as the player reaches half of the level goal, the tempo of the background music will be slightly increased. Moreover, one quarter away from the goal music noticeably speeds up to distract the player and thereby suborn him/her to make quick decisions.

Level	Elements varying cloud estimations
1st	Player's arousal level is presented but it has no effect on the game at all
2nd	Player's arousal level affects the game
3rd	Estimation clouds move simulating the wind
4th	Estimation clouds become bigger and smaller in sizes
5th	Estimation clouds are same sized, but fake clouds with text start to appear
6th	Fake clouds with numbers start to appear
7th	Estimation clouds become bigger and smaller in sizes again
8th	Player has to achieve targeted arousal, but fake clouds do not appear in this level
9th	Fake clouds with numbers start to appear again
10th	Estimation clouds' speed and appearance time of the clouds are random
11th	Entering Death Mode, game speeds up impossibly

Table 6.1: Game level elements varying cloud estimations.

  - The speed of cloud appearance increases while the time for decision decreases. As the player progresses through a level, the cloud estimations slightly increase their movement speed; moreover appearance time and decision time slightly decrease to distract the player into making quick decisions in an attempt to make him commit errors.

- **Game elements affected by arousal**
  - Distribution of price signals is dependent on the arousal level of the player. The further s/he is away from the target arousal, the larger will be the spread of estimations. This will make it more difficult to calculate the true price.
  - Distribution of true price in the next period is dependent on the arousal level of the player. Every round a good shifts its true price on the market. The further s/he is away from the desired level of arousal, the larger deviation of the next true price is. This will make true price shift more unexpectedly.
  - Speed of cloud movement is directly linked to arousal. As the game progresses clouds start moving. The further s/he is away from the desired level of arousal, the faster the movement of clouds becomes. This makes it harder to visually observe the price estimation.

To keep the game interesting, piecewise elements varying cloud estimations are presented through the levels (Table 6.1). Note that every element adds to all the active ones from previous levels.

6.5 Evaluation

The Auction Game has been thoroughly tested using functionality, heuristics and play testing.

6.5.1 Functionality and Heuristic Evaluation

The functionality evaluation consists of three categories: validation, verification, and future support. The validation was done in collaboration between the partners from xDelia at several face–to–face meeting where the product was demonstrated. By letting the product owner test it before it was finalized much of the rather abstract discussions became concrete as feedback was communicated.

The verification was done by the development team throughout the development process. By working in an incremental development style the product was constantly tested for errors when new things were added, thus minimizing latent errors. The future support is achieved by designing the game around a modular and dynamic architecture, making it easy to adjust for future studies; moreover, a xAffect as separate component is used for measurement of the arousal allowing for other sensors to be used within the same game.

Development of the Auction Game was followed by a heuristic evaluation (Isbister and Schaffer, 2008) aiming at qualitatively identifying design errors and suggest improvements. A group of expert evaluators reviewed the game using heuristics divided into a set of categories, looking for potential usability and gameplay problems in the prototype. Heuristic evaluation pinpointed several important design issues and reinforced the robustness of the game.

6.5.2 Playtesting Evaluation Method

A total of six students volunteered to participate in the Playtesting Evaluation. They were all students of *Blekinge Institute of Technology* aged between 20 and 32 years old with four of them being male and two female. They reported varying gaming experience.

Before the game, the students were fitted with the Movisens ecgMove HR sensor and given a tutorial session. In order to objectively determine which game elements the players were paying attention to, the game was played through the Tobii T60 eyetracker logging data on different Areas of Interest (AOI) and recording the whole gaming session on video. The purpose was to be able to tell how important different visual objects (AOIs) on screen are to a player. At the end of each game, each participant was given: an Emotion–Regulation Questionnaire (Gross and John, 2003) in order to identify suppression and reappraisal tendencies of individuals; a modified System Usability Scale (Nacke et al., 2010; Brooke, 1996) questionnaire measuring game usability. The questionnaire contained 10 questions whose score was summed up in a single number representing a composite measure of the overall usability of the game being studied; an interview session where participants could openly discuss perceived game speed and difficulty, as well as visual cue elements and any issue they wanted to note.

6.5.3 Playtesting Evaluation Results

The Auction Game scored a mean value of 67.92 in a range from 0 to 100 on modified SUS questionnaire. Thus according to (Albert and Tullis, 2013) where a score of 60 presents a border between poor and average usability, we can conclude that the game fulfills the average game usability.

The game was successfully played up to a hard 8th level by two participants, both of which are high reappraisers, while one had low and the other normal suppression tendencies. They both evaluated the game as manageable and in the interview session reported that they were practicing emotion–regulation techniques themselves without being instructed at all. This provides evidence for a good game design of the Auction Game.

Five out of six participants reported that they were not paying attention at all to the arousal meter indicator present at the top–right of the screen. We evaluated this claim on how informed the participants were about their arousal level. They have an option of keeping track of it on the arousal meter indicator during rounds in the whole gaming session. A paired–samples t–test was concluded on the eyetracker data to evaluate the difference in number of gaze observations on marked indicator arousal meter AOI compared to number of rounds taken for each participant. There was a statistical significant difference found with number of rounds (M = 110.17, SD = 95.26) to number of arousal meter observations (M = 16.33, SD = 24.5, $t(5) = 2.94$, $p < .05$). Thus we can say that participants paid little or no attention on the arousal meter indicator during the whole playing session. Participants reported that the reason for paying little or no attention on the arousal meter indicator was lack of time during fast paced decisions. Most of the participants reported that they were paying attention to the arousal indicated by the color of the cloud estimations, especially when it turned red. This gave evidence to concentrate on making the color of the cloud estimations more distinct, since players are focusing their concentration on them. Further studies should identify how to optimally present the arousal information to the cognitively engaged player during fast paced decisions.

A one–way between–groups analysis of variances was conducted to explore the impact of arousal level on profit in each round. Total number rounds played, 661 rounds, were divided into 5 groups according to the arousal level while decision was made (Group 1: 1[relaxed], Group 2: 2 . . . Group 5: 5[highly aroused]). There was a statistically significant difference at the $p < .05$ level in profit made each round for the five arousal groups [F(4, 656) = 3.566, $p < .01$]. The effect size, calculated using eta squared was .02. Post–hoc comparison using Turkey HSD test indicated that the mean score for Group 1 (M = .6328, SD = 2.95) was significantly different from Group 5 (M = -1.369, SD = 3.3). Other groups did not differ significantly. Same has been conducted for the time needed to reach a decision in seconds and there was a statistically significant difference the $p < .05$ level [F(4, 656) = 5.753, $p < .001$] between Group 5 (M = 1.55, SD = .45) and rest of the groups. The effect size, calculated using eta squared was .03. This gives strong evidence supporting a good design of the Auction Game to present a hard challenge and punishment to a player in an undesirable high arousal emotional state.

6.6 Discussion and Conclusion

Evidence shows that emotions impact our decisions, especially in the field of finance. Thus it makes sense to develop a tool to get people aware of this implication as well as to help them in regulating their emotions to reach better financial decisions. A serious game emerged as an appropriate tool in which players get feedback on their emotional arousal, according to their psychophysiological state. This on screen feedback helps subjects to get aware of and to learn how to control their emotional state. The Auction Game is a serious game where a player buys or sells stocks with the objective to train emotion–regulation; but also to get them aware of the arising emotions. To support this, achieving a target arousal level will reward the player accordingly, increasing his/her earned profit.

Data from successful participants gives first evidence that the Auction Game is indeed overwhelming and puts players at a highly aroused state where they need to practice emotion–regulation techniques to succeed in the game.

We have demonstrated in the Auction Game how one can reward a player achieving a desired arousal level, while at the same time presenting a hard challenge and punishment to a player in an undesirable high arousal emotional state. Through this experience emotion–regulation can be learned and practiced using this tool.

For future work it is planned to use the Auction Game in varying contexts: it will be interesting to detect which strategies players apply (e.g. breathing) in order to regulate their level of emotional arousal; Related to this question, we want to examine how effective certain strategies will prove, measured either by self–perception and/or by physiological measures (e.g. phasic heart rate response, heart rate variability), in order to perform well in the game. Moreover we want to find out whether certain emotion–regulation strategies (suppression versus reappraisal) result in systematic differences in game performance. Our last and most prominent goal is to evaluate the Auction Game as a learning tool for enhanced emotion–regulation, i.e. examining whether extensive playing of the Auction Game (or another tool following the same paradigm) can systematically improve subjects skills to get aware and control their emotions and whether these skills are transferable to other (financial) tasks, leading to a long lasting shift in decision performance. This future research has to be conducted in order to investigate how successful our approach is in teaching emotion–regulation and how well it can be transferred to real life trading. Up to now, we have demonstrated that the Auction Game was successful at reaching its goals as a study tool, as well as a usable game. If we can systematically succeed in this, we can make learning emotion–regulation in the context of financial decision–making more fun and more effective.

Chapter 7

Integrating Biosignals into Information Systems: A NeuroIs Tool for Improving Emotion Regulation

Abstract

*Traders and private investors are aware that emotional processes can have material consequences on their financial decision performance. However, typical learning approaches for de-biasing fail to overcome emotionally driven financial dispositions; mostly because of subjects' limited capacity for self-monitoring. Our research aims at improving decision makers' performance by (i) boosting their awareness to their emotional state and (ii) improving their skills for effective emotion–regulation. To that end we design, implement and evaluate a serious game based Neuro*IS *tool that continuously displays the player's individual emotional state, via biofeedback, and adapts the difficulty of the decision environment to this emotional state. The design artifact is then evaluated in two laboratory experiments. Taken together, our study demonstrates how information systems design science research can contribute to improving financial decision–making by integrating tools from cognitive neuroscience in information technology artifacts. Moreover, we provide specific design guidelines for how biofeedback can be integrated into information systems.*

7.1 Introduction

The human nervous system is marvelous. While in one moment it enables us to achieve top performances—be it intellectual or physical—the next moment it allows us to relax and enjoy the beauty of a quiet Sunday afternoon. With and without conscious awareness, the human nervous system thereby continuously adjusts visceral functions in order to match the entire system to ongoing changes in environmental demands. A core element of this process is emotion. Emotions facilitate our interactions with the socioeconomic environment; they foster meaningful interpersonal interactions, prepare behavioral responses, and enable us to take advantageous decisions (Gross, 2009). Unfortunately, however, emotions are not always accurately processed. Emotions can get “out of control” (Goldstein, 1996). In the heat of the moment, human decision makers can be overwhelmed by their emotions and—under the influence of high levels of arousal—lose control over their actions. In business, such processes can have material consequences for the organizations the decision makers represent.

Unfortunately, there is a strong misbalance between emotions and their consequences. While the emotions are transient, the consequences of the actions are not; in fact, they can be long-lasting. This applies specifically for important decisions, because “important decisions induce powerful emotions in decision makers” (Loewenstein, 2000, p. 429). In other words, it is vital for decision makers and for the organizations they represent to adequately process emotions, because emotions particularly affect important decisions. Financial decisions fall into this category (Lucey and Dowling, 2005). There are, however, strong interpersonal differences in the capabilities of adequate emotional processing (Lo and Repin, 2002; Lo et al., 2005). It has been shown that traders and investors with high emotion–regulation capabilities perform better in trading (Fenton-O’Creevy et al., 2012b). In contrast, decision makers with low emotion–regulation capabilities frequently are more prone taking disadvantageous decisions with undesirable consequences. Thus, improving the decision makers’ emotion–regulation capabilities is an important issue for their organizations and there is a need for sophisticated Is tools that provide decision support and facilitate the skill development process.

In this paper, it is our objective to help decision makers with improving their emotion–regulation capabilities in the context of financial decision–making. Therefore, we design, implement, and evaluate a NeuroIs tool (vom Brocke et al., 2012). More specifically, we provide a dynamic learning environmentthat is based on a serious game with real-time biofeedback. Biofeedback aims at providing subjects with information on their own physiological processes, such as by measuring skin conductance or heart rate, in order to improve performance, health, or user experience (Nacke et al., 2011; Varvogli and Darviri, 2011). In our game, the player faces a financial decision scenario that is directly linked to his or her individual emotional state which is also indicated via biofeedback. Based on heart rate measurements, the game incorporates the player's arousal level into the game task. The player wears an electrocardiogram sensor that transfers the data via Bluetooth to the game. Depending on the player's ability to regulate emotions, the financial decision scenario of the game continuously adjusts and thereby becomes more (or less) difficult. Our design artifact is thus also a use case for how information systems can incorporate physiological data (vom Brocke et al., 2012).

7.2 Conceptual Foundations

7.2.1 Emotion and Cognition in Financial decision–making

Classical economic theory has "idealized the decision maker as a perfectly rational cognitive machine" (Sanfey et al., 2003, p. 1755). This consequentialist perspective, however, leaves no space for an influence of emotions on cognition and decision–making, and struggles with a broad body of literature which provides ample evidence that the assumption of a perfectly rational homo economicus does not hold for most human decision makers (Bechara and Damasio, 2005; Shiv et al., 2005). From an economic point of view, the occurrence of emotions was originally considered as a violation of expected utility theory and their influence as disturbing or even counterproductive for financial decision–making (Bechara and Damasio, 2005). With respect to financial decisions, emotions are suspected to be the underlying driving force for a whole group of biases, such as the disposition effect (Weber and Camerer, 1998), loss aversion (Heilman et al., 2010; Sütterlin et al., 2011), and the phenomenon of auction fever (Adam et al., 2011; Ku et al., 2005). These biases have been shown in a broad range of field studies (Ku et al., 2005; Shefrin and Statman, 1985), and can result in considerable financial losses and therefore have material consequences for the stakeholders.

While emotions can have biasing effects, insights from economic psychology widened the picture. In this regard, the majority of psychologists now view decision–making in a dual process framework, stating that depending on the context, humans process reality in two fundamentally different ways: the slow and analytical rational system and the fast and intuitive experiential system (Kahneman and Frederick, 2002). Especially in environments where fast processing is necessary, subjects tend to rely on their emotions and follow an "affect heuristic" (Slovic et al., 2007). Essentially, the affect heuristic states that human decision makers integrate context specific affective feelings into perception and their decision–making process. Research in cognitive neuroscience provided evidence that activation in distinct neural circuits, correlating with positive as well as negative affective states, promote and avoid varying types of risky financial mistakes (Kuhnen and Knutson, 2005; Peterson, 2007). In this regard, Kuhnen and Knutson concluded that “financial decision–making may require a delicate balance” and that “excessive activation of one mechanism or the other may lead to mistakes” (Kuhnen and Knutson, 2005, p. 767).

The fact that emotions have an impact on financial decision–making became evident in a variety of studies (e.g. Adam et al., 2013; Goldstein, 1996; Riedl et al., 2010). It is important to note, however, that most of these studies refer to contextualized situations and “the context can influence how we process stimuli that may have affective properties” (Rolls and Grabenhorst, 2008, p. 230). In other words, affective processes are context dependent and at this stage it is not clear what in general the emotional ingredients are that guide optimal financial decision–making. Bechara specifically accounted for the interplay of emotions and information processing—and found that emotions in fact can contain relevant information (Bechara, 2004). In their somatic marker hypothesis, Bechara and Damasio concluded that emotions seem to be both a source of biases as well as an important mechanism for advantageous decision–making (Bechara and Damasio, 2005). The authors showed that the experience of emotions itself is a mandatory prerequisite for advantageous decision–making. More specifically, the authors found that those subjects with brain lesions, that are critical for the processing of emotions, performed worse in a card game under ambiguity than healthy subjects (Bechara and Damasio, 2005; Bechara et al., 1997). Bechara and Damasio concluded that so called “somatic markers”—essentially emotional responses to information events—beneficially guide our focus of attention. The somatic marker theory hence states that “decisions are aided by emotions, in the form of bodily states, that are elicited during the deliberation of future consequences and that mark different options for behavior as being advantageous or disadvantageous” (Naqvi et al., 2006, p. 260). In other contexts, however, it was shown that such emotional processes can also have disruptive effects (Shiv et al., 2005). It is therefore important to note that emotions cannot always “be trusted as leading to good or bad decisions” (Shiv et al., 2005, p. 438). Still, a well-functioning recognition of these emotional bodily states is a necessary prerequisite to be equipped with the skills to decide whether it is better to follow an emotional response or to inhibit it.

Traders and investors are well aware of the influence emotions can have on their decision performance, and are therefore highly interested in eradicating such emotional dispositions (Fenton-O’Creevy et al., 2011). In an interview–based study with professional traders Fenton-O’Creevy et al. (2011) found that major losses often resulted in periods of risk aversion and overcautious behavior, whereas major gains resulted in excessive self-confidence and unwary behavior. Similar results were found in laboratory experiments (Novemsky and Kahneman, 2005). Traders and private investors have already realized that their emotional states cannot simply be ignored, and that certain decision–making biases, such as for instance illusion of control (Fenton-O'Creevy et al., 2003), are even beyond conscious awareness.

In this context, Lo and Repin (2002) conducted physiological measurements in a study with professional day traders and found that traders expose high arousal during the trading day, with strongest activation during periods of high market volatility. In a subsequent study, Lo et al. found that "extreme emotional responses are apparently counterproductive from the perspective of trading and performance" (Lo et al., 2005, p. 357). This provides further evidence that arousal can in fact significantly threaten rational cognition and decision–making (Kuhnen and Knutson, 2005).

Poor impulse control strategies, high emotional reactivity and the exposure of high levels of arousal, which might threaten rationality, are often associated with poor investment performance (Peterson, 2007). It has also been shown that trading success is also correlated with personality traits. In particular, introversion, emotional stability, and openness to new experiences have been found to be associated with good trading performance (Peterson, 2007). Peterson found that it is important for traders to be aware of their own "fallibility" and to avoid to be overpowered by their emotions. The author recommended traders to "use awareness of [their] emotional state to generate a personal warning signal" (Peterson, 2007, p. 76). Likewise, Biasis et al. (Biais et al., 2005, p. 308) found that self-monitoring can enhance trading performance. In this vein, an increased awareness to one's own emotional state can help traders to identify a state fueled with high arousal (Fenton-O'Creevy et al., 2012a). In addition to the mere awareness of arousal, however, it is also important to develop strategies in order to "interpret and manage affective states" (Peterson, 2007, p. 75).

7.2.2 Emotion–Regulation

In this context, it is important to introduce emotion–regulation (Er)—a concept that has drawn much attention in psychology in recent years (Gross, 2009). Following the concept of emotion–regulation, emotions can determine and alter the way subjects perceive a certain situation and thereby influence the way they react to it. Thus, emotions act as response tendencies: they suggest distinctive responses to a certain situation, which an individual may or may not follow (Gross, 1998b). In this sense, emotions "reflect the status of one's ongoing adjustment to constantly changing environmental demands" (Thayer et al., 2009, p. 85). The process model of Gross (Gross, 1998a) is widely known and acknowledged in the field of Er research.[1] Gross's process model stems from the assumption that emotions are generated in an emotion generative process. Thereby, research distinguishes primarily between antecedent- and response-focused Er strategies. Antecedent Er strategies apply while the emotion is still unfolding and has not reached its peak. The most prominent representative is cognitive reappraisal.

Consider, for instance, an impatient It manager waiting for a developer to fix a potential security hazard in the system. Cognitive reappraisal could now result in becoming aware that the waiting time could be spent more usefully, for instance to reassure that the system was recently backed up. On the other hand, response-focused Er aims at altering and controlling the experiential, behavioral, and physiological response when the emotion has already unfolded completely. The most prominent strategy for response-focused Er is suppression. In the It manager example, suppression could enfold in the way that the manager gets really irritated but then decides to not make the situation even worse by expressing this emotion. In fact, emotions and their regulation can be both conscious- and subconsciously processed (Gross, 2009; Williams and Bargh, 2009). Therefore, an increased emotional awareness is crucial for enhanced Er. The applied Er strategies seem to be determined by personality traits and individual psychological capacity (Gross, 1998a).

7.2.3 Improved Financial decision–making through Enhanced Emotion–Regulation

Previous research has shown that effective Er can improve financial decision–making. Heilman et al. (2010) and colleagues showed that Er can reduce loss aversion and thereby contribute to advantageous decision–making. Sokol-Hessner et al. (2009) induced an intentional cognitive regulation strategy which emphasized "perspective taking". The authors showed that applying this form of cognitive reappraisal can mitigate physiological responses to losses relative to gains and also reduce loss aversion. Sütterlin et al. (2011) measured subjects' individual degree of inhibitory control and found it to be negatively correlated with sus-

[1] Previous to the Er definition by Gross (Gross, 1998a), similar concepts have been introduced, e.g. on coping, mood regulation, and traditional ego-defenses. Er, however, has been clearly distinguished from them (Gross, 1998b). Other approaches from clinical psychology also cover additional Er techniques such as acceptance, avoidance, problem-solving, and rumination (Aldao et al., 2010). When comparing the number of citations, Gross's model is the most prominent at this stage.

ceptibility to framing effects in a risky-choice task. Similarly, Fenton-O'Creevy and colleagues measured the heart rate variability of professional traders in their daily environment and found a significant relationship between effective Er and trader experience. In an interview study, Fenton-O'Creevy et al. (2012b) found that high–performing traders not only have a better awareness of their current emotional state but also are more sophisticated in down–regulating high levels of arousal. The authors reported that inexperienced traders simply try to avoid negative emotions, while more experienced traders use different strategies and do not evade these negative unfolding emotions.

In summary, previous research has shown that the awareness of one's emotional state and the capacity to regulate levels of high arousal overall is an important attribute for financial decision–making. Er can help to mitigate emotion related decision biases and eventually lead to better decision performance. However, how can a decision maker actually learn to improve his or her emotion–regulation capabilities? Unfortunately, traditional learning approaches seem to fail when it comes to improving emotion–regulation capabilities (Fenton-O'Creevy et al., 2012a). In the context of Er, common problems of traditional learning approaches are (1) that learning Er and actually experiencing high levels of arousal are separated in time, and (2) that the subjects can get practically no feedback on their own level of Er. Research in cognitive neuroscience and psychophysiology, however, found that Er has an influence on the underlying visceral processes of emotional experience (Martin and Delgado, 2011). Applying such methods may therefore be a promising approach for improving emotion–regulation capabilities, because in this vein the users can be provided with direct feedback on their emotional state.

7.2.4 NeuroIs and Design Science Research

The nascent field of NeuroIS is drawing upon the theories, methods, and tools offered by cognitive neuroscience and psychophysiology (Dimoka et al., 2011, p. 687) (Riedl et al., 2009). For instance, Riedl et al. (2011) showed that deciding whether to trust an avatar induces less intense neurobiological processes than deciding whether to trust an actual person. In a different study, Riedl et al. (2010) reported that there are gender differences with respect to the activity of specific brain areas when deciding on the trustworthiness of eBay offers. Based on heart rate measurements, Adam et al. (2013) showed that in electronic Dutch auctions the impact of time pressure on final prices is mediated by arousal.

While there is a growing number of empirical NeuroIs papers, the consideration of cognitive neuroscience "is still in its infancy in Is design science research" (vom Brocke et al., 2012, p. 2). In order to provide a framework for how tools and theories from neuroscience can be applied in IS design science research, vom Brocke et al. (2012) derived three specific applications strategies (Strategy 1: inform the building and evaluation of IT artifacts; Strategy 2: use of neuroscience tools to evaluate IT artifacts; Strategy 3: use neuroscience tools as built-in functions of IT artifacts). In this paper, we focus specifically on how neuroscience tools can be used as built-in functions of IT artifacts (see Strategy 3 in the framework of vom Brocke et al. (2012)). More specifically, vom Brocke and colleagues argued that "IT artifacts with built-in neuroscience tools may even adjust to the affective state of the user" (vom Brocke et al., 2012, p. 9). Thereby, IS design science research can build on advances in the fields of affective computing (Nacke et al., 2011; Picard, 2000) and neuroergonomics (Di Stasi et al., 2011; Parasuraman, 2003).

Affective computing is defined as "computing that relates to, arises from, or deliberately influences emotions" (Picard, 2000, p. 3). In this regard, Picard and colleagues stressed the importance of software intelligence having the ability to sense and respond to the users' affective states based on physiological measurements (Picard et al., 2001). Neuroergonomics is defined as "the study of brain and behavior at work" (Parasuraman, 2003, p. 5). In this context, Di Stasi and colleagues showed how eye-tracking devices can be used to monitor operators' mental workload when interacting with hypermedia (Di Stasi et al., 2011). Moreover, previous research used physiological measures to investigate how interacting with information technology can induce considerable levels of stress in the users (Riedl et al., 2012; Zhai and Barreto, 2006b). A core concept in affective computing and also in neuroergonomics is biofeedback. Biofeedback aims at displaying physiological parameters such as heart rate or skin conductance acoustically or visually (Lehrer et al., 2000; Nacke et al., 2011; Ouwerkerk, 2011). By this illustration, the subjects get direct feedback on their own visceral processes; processes which are for most subjects usually almost unperceivable (Dawson et al., 2011). Biofeedback can increase the user's attention to his or her emotional state, which in turn can improve interoception, i.e. the conscious awareness of one's own physiological processes (Vaitl, 1996). Previous research has shown that biofeedback can help to successfully reduce heart diseases and depression (Lehrer et al., 2000). In the context of financial decision–making, the electrics company Philips together with the Dutch bank ABN AMRO developed a biofeedback device for retail investors. The "Rationalizer" is a device that "acts as a kind of emotion mirror in which the user sees reflected the intensity of his feelings in form of dynamic lighting patterns" (Djajadiningrat et al., 2009, p. 39).

7.3 Design

7.3.1 Requirements

To help users improve their emotion–regulation capabilities, we designed, implemented, and evaluated a NeuroIS tool. In the following, we define a set of specific requirements (R1–R3), which are based on the literature and serve as guidelines for the design process. Later, these requirements are used as criteria for the evaluation of the design artifact (Peffers et al., 2007).

The first requirement (R1) refers to the general nature of the learning environment we seek to create. Research in educational psychology has shown that motivation, positive affect, and interactivity have a positive influence on learning (Dror, 2008). In order to promote learning performance, it is therefore important to provide learners with engaging learning environments that enable them to actively interact with the material (Dror, 2008). Moreover, down regulation of high levels of arousal can only be actively practiced, when the learners actually face an arousing environment. Thus, in order to create a design artifact based upon which ER can be actively practiced, the users actually have to face an engaging learning environment which can elicit arousal. Finally, the general focus of our paper lies on improving ER in the context of financial decision–making. In order to facilitate learning transfer from the game into real-life contexts, the learning environment should also be embedded in a financial context.

Requirement 1 (R1): *The design artifact has to provide an engaging learning environment which can elicit arousal and is embedded in the context of financial decision–making.*

The second requirement (R2) refers to our overall goal of helping users with improving their emotion–regulation capabilities. Previous research has shown that providing learners with immediate feedback and reward can promote positive affect and motivation, which in turn facilitates the learning process (Annetta, 2010). This particularly applies when subjects have difficulties with assessing their own performance (see R3). By directly rewarding effective ER and punishing poor ER, the design artifact should thus help users to actively learn ER.

Requirement 2 (R2): *The design artifact has to provide an environment in which effective* ER *is rewarded.*

The third requirement (R3) refers to the biofeedback presented to the user. A major problem when improving emotion–regulation capabilities is that learners can get practically no feedback on their own engagement with ER (Fenton-O'Creevy et al., 2012a). Therefore, we aim at providing the users with real-time biofeedback on their emotional state by means of physiological measurements. In this vein, we intend to boost users' awareness to their own physiological state and thus improve their interoception. This live biofeedback has to meet specific requirements. First, from a technical perspective, biofeedback systems, which are intended to be used in daily life, should generally be designed to be "unobtrusive and as unnoticeable as possible" (Ouwerkerk, 2011, p. 24). In our project, the game was intended to be used in day-trading centers. Therefore the device should be easy and unnoticeable to wear for traders and investors. Second, from a user-interface perspective, the biofeedback should be provided in a meaningful way. In other words, the users should be able to actually draw useful information from the displayed biofeedback. Taken together, by providing the user with live biofeedback in an unobtrusive and meaningful way, the users can get direct feedback on their individual level of Er.

Requirement 3 (R3): *The design artifact has to provide the user with live biofeedback in an unobtrusive and meaningful way.*

7.3.2 Concept of the Learning Environment

With respect to requirement R1, previous research has shown that serious game based learning is engaging and has positive effects on the development of new knowledge and skills (Anderson and Lawton, 2009). In general, serious games can be understood as games which aim at purposes other than entertainment alone (Corti, 2006; Léger et al., 2011). Repeatable, highly engaging, and motivating serious games can result in improving training activities and therefore in enhanced skill development (Corti, 2006). One reason for the effectiveness of serious games as learning instruments is that "participants in a serious game must try something, even though they may not have complete or clear information about the best course of action" (Léger et al., 2011, p. 45). Due to these characteristics, we decided to design and implement our design artifact as a serious game. The game is titled Auction Game and, in line with R1, is embedded in a financial context. The goal of the player is to keep his or her own level of arousal low.

In the Auction Game, the player is given the role of a trader who has to earn money by buying and selling a single stock in consecutive trading rounds. In each trading round, the player is provided with three sequentially displayed price estimates for the real value of the stock in this round. The player is informed that these estimates correspond to the ratings of three independent analysts and that the true value of the stock always equals exactly to the mean of the three price estimates. Thereby, we follow the economic concept of common value signals (e.g.

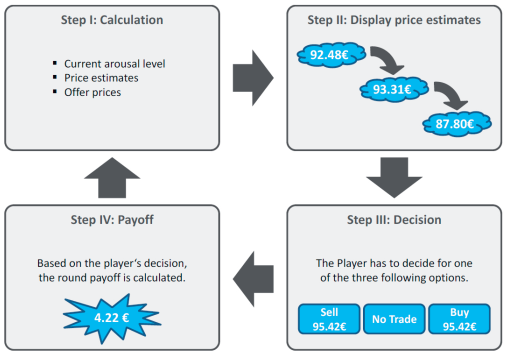

Figure 7.1: Process of one trading round.

Oh, 2002). The player's task in each round is to calculate the mean of the three price estimates and then decide whether to buy or sell the stock at a given price. If the price is below (above) the true value of the stock, the player can make a profit by buying (selling) the stock.[2] If the player opts for the wrong option, the payoff of this round is negative.

Figure 7.1 depicts the process of a single trading round. Each trading round comprises four steps (Step I–IV). In Step I, the game calculates the user's current level of arousal, the three price estimates for the stock, the true value of the stock, and an offer price. Later the player can buy or sell a stock for that offer price. In Step II, the three price estimates are consecutively displayed to the user. In the example in Figure 7.1, the price estimates are € 92.48, € 93.31, and € 87.80. Thus, the true value of the stock is (€ 92.48+€ 93.31+€ 87.80)/3 = € 91.20 in this round. In Step III, the user has to decide for one of three options. In the example of Figure 7.1, the user can buy € 95.42, sell for € 95.42, or press the no-trade button. By pressing the so called no-trade button, the user can jump to the next round without having to suffer from a negative payoff. Conceptually, the no-trade button enables the user to escape an immediate decision and take a step back to calm down before returning to the active trade. If the user does not choose one of the three options within two seconds, the payoff is € –5.00. This is by design the worst payoff a player can achieve in a single round. Therefore, action should always be preferred to no action. In Step IV, the payoff of this round is displayed. In the example of Figure 7.1, the player decided to sell for € 95.42. Since the mean of the price estimates (€ 91.20) is lower than the offered price (€ 95.42), the player makes a profit of € 4.22. The game then continues with Step I of the next trading round.

[2]Short selling here means that the player can sell the stock without actually having it. This is a common procedure in financial markets.

7.3.3 Rewarding Emotion–Regulation

In order to meet R2, the Auction Game is designed to reward effective Er based on physiological measurements. Thereby, we focus specifically on down-regulation of high levels of arousal as such levels are often found to be correlated with disadvantageous financial decision–making (Adam et al., 2013; Ku et al., 2005; Kuhnen and Knutson, 2005). It is important to highlight, however, that our approach is not normative in the sense that we aim at dictating specific actions or strategies. To the contrary, it is our goal to extend decision makers' adaptive toolbox (Gigerenzer, 2002) by improving their emotion–regulation capabilities and thus providing them with the ability to perform Er when needed.

To reward down-regulation of arousal, the Auction Game is designed such that a user's current level of arousal (on a normalized scale from 1 to 5) has an influence on the variance of price estimates. Therefore, by design, the variance of price estimates increases with the current level of arousal. For instance, for an arousal level of 1, the price estimates would be € 92.48, € 93.31, and € 87.80. For an arousal of 5, however, the variance of price estimates is much larger, which results in price estimates of € 68.22, € 79.21, and € 126.17. Thus, the game makes it more difficult for the user to compute the mean of the price estimates while at the same time the user cannot gain any advantage from larger variance in the price estimates. Based on this design, we directly connect the user's current level of arousal to game difficulty. In this way, the game not only helps to increase emotional awareness, but even directly rewards down-regulation of arousal. In later levels, other elements such as cloud movement on the screen or the variance of price gaps[3] between consecutive decisions increases with arousal. Therefore, the better a subject is able to down-regulate arousal, the easier the financial decision scenario of the game becomes. In summary, the Auction Game is designed to directly reward effective ER in the context of financial decision–making.

7.3.4 Arousing Game Elements

The Auction Game incorporates specific design elements that aim at inducing arousal in the users (R1). First, the Auction Game is designed to continuously elicit immediate emotions in the users. Previous research has shown that being confronted with the positive and negative outcomes of one's own financial decisions can induce rewarding and aversive emotions (Astor et al., 2013a; Bechara and Damasio, 2005). Therefore, every time the user has submitted a decision,

[3] It is important to note that subjects cannot benefit from the price gap or the price variance. These elements only make it more difficult to predict the general price level in the next round.

the Auction Game immediately provides a visually and acoustically highlighted feedback on whether the user has achieved a gain or a loss. Second, the Auction Game is designed to increase the general level of arousal by constantly putting the users under time pressure. Time pressure is generally known for increasing the decision makers' arousal and their willingness to take risks (Adam et al., 2013; Ku et al., 2005). In particular, in each trading round the user has only 2 seconds to submit a decision. Similarly, each estimation cloud is displayed for only 1 second, and this display time reduces by 25 milliseconds with each new level. Taken together, the duration of one round is the time needed for the cloud display plus 2 seconds for submitting a decision.

To induce another level of time pressure, and also to make the game more engaging (Kelly et al., 2007; Annetta, 2010), the Auction Game is structured into seven game levels. The goal of the user is to reach and successfully complete the highest level. To advance from one game level to the next, the user has to make profitable decisions to achieve the current game level's profit goal. In the evaluation studies, the users needed, on average, 22.38 decisions to reach the next level. In every new level, the profit goal increases by € 30.00. When the user falls € 10.00 below the starting point of the current level, the game ends. The game is conceptualized such that the user either successfully completes level 7 or has to quit the game because of time or money constraints. Optimally the game should run for approximately 25 minutes with a total profit of around € 200.00 in the game.

In order to induce another level of time pressure, the user has only four minutes to reach the goal of each game level. Moreover, the tempo of background music in the game automatically adjusts to the user's current profit level. Research in psychophysiology has shown that the tempo of background music has an impact on arousal (Critchley et al., 2004). Therefore, when the player has achieved 50% of the level goal, the tempo of the background music, slightly increases (2% total increase in pitch). The original tempo is 120 bpm. When the player has achieved 75% of the level goal, the background music noticeably speeds up again (5% total increase in pitch).

Level	**Elements varying cloud estimations**
1/7	Player's arousal is displayed but it has no effect on the game at all.
2/7	Player's arousal level affects the variance of estimates.
3/7	Estimation clouds move.
4/7	Estimation clouds become bigger and smaller in size.
5/7	Estimation clouds are same sized, but fake clouds with text start to appear.
6/7	Fake clouds with numbers start to appear. Distracting background images.
7/7	Estimation clouds become bigger and smaller in sizes again.

Table 7.1: Overview of game levels.

The seven different levels of the game are summarized in Table 7.1. As the game progresses the game difficulty changes, making it more difficult for the player to take a profitable decision. Moreover, from level to level, more and more distracting elements are included in the game environment. Among those are additional irrelevant clouds carrying false information, reduced display time of the clouds, as well as auditory and visual distracting elements. For instance, starting in level 6, the background of the game interface shows arousing photos of the International Affective Picture System (IAPS) database (Lang, 1995). Also, the speed of cloud movement is correlated to arousal in higher levels. Another reason for including distracting elements in the game is that in demonstration sessions, practitioners from the field reported that traders and investors continuously have to deal with arousing distractions during trading.

7.3.5 Biofeedback

A central element of the Auction Game is biofeedback. Research in psychophysiology has shown that "the experience of both emotions and stress are known to be accompanied by a physiological state of arousal" (Grandey, 2000, p. 99), which in turn manifests in changes in physiology. In the context of IS, Riedl et al. (2012) have shown that the stress hormone cortisol is released when users experience "technostress" in response to a system breakdown. While measuring cortisol levels is undoubtedly highly reliable for assessing arousal, the use of saliva samples is not suited for providing users with a continuously adjusting live biofeedback (Ouwerkerk, 2011).

In order to meet R3 and integrate live biofeedback into the Auction Game, we need a real-time measure that accounts for the following contextual and situational circumstances of the user. First, the Auction Game is intended to be used by traders and investors in their daily environment; therefore, physiological measures should be as unobtrusive as possible. Second, the Auction Game is, by design, characterized by a fast-paced decision environment, and affective processes should thus be reflected in the measure as quickly as possible. Third, the degree of interplay between the sympathetic nervous system (fight or flight) and the parasympathetic nervous system (recreation) provides an indication for a subject's emotion–regulation capabilities (Sütterlin et al., 2011). In order to employ a measure that reflects the activity of both branches of the nervous system, we decided to use heart rate as the underlying parameter for biofeedback. Sympathetic and parasympathetic activation is almost instantly reflected in changes in heart rate (Berntson et al., 2007). Moreover, the heart rate can be measured in an unobtrusive way by use of dry electrodes in a chest strap or a shirt. This allows a movement of the hand and will neither cause artifacts nor loose electrodes.

7.4 Implementation

The implementation of the Auction Game followed an iterative process (Peffers et al., 2007) which involved demonstration to real practitioners at trader and investor shows, as well as play-testing calibration with student testers.

7.4.1 Game Architecture

The Auction Game is based on a modular and dynamic architecture, which can be adjusted to include further game elements by modifying an XML (Extensible Markup Language) configuration file. In particular, the Auction Game can be used with measurement devices from different manufacturers. A screenshot of the game and further details on the game mechanics are provided in the Appendix.

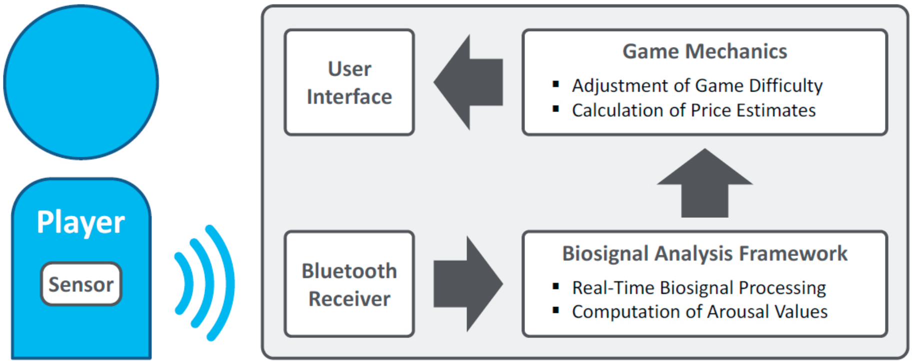

Figure 7.2: Schematic overview of the game architecture.

The Auction Game architecture is schematically depicted in Figure 7.2. The Ecg device was chosen in consideration of R3 to be as unobtrusive as possible. More specifically, the player wears a dry electrode chest-belt and a sensor that transmits the user's electrocardiogram (Ecg) wirelessly to a Bluetooth receiver. A biosignal analysis framework then processes the Ecg data, computes the arousal values and feeds them into the game in real-time. The game constantly adjusts the level of game difficulty, calculates the price estimates, and updates the user interface.

7.4.2 Computation of Arousal Values

As outlined in the design section, the computation of arousal values is based on heart rate. To connect the physiological sensors to the game, the xAffect framework was used. This software environment provides a modular open source framework for offline and online analysis of biosignals. After receiving the Ecg signal, Qrs detection is performed using the Osea algorithm (Hamilton, 2002). The Qrs algorithm is used to detect these R peaks, based on which heart rate is computed.

Before the actual trading task starts, there is a five minute rest period in order to assess an individual baseline level of arousal (cf. Sütterlin et al., 2011). The arousal values in the game are then derived from the heart rate in relation to this baseline period. Sympathetic activity tends to increase heart rate, while parasympathetic activity decreases heart rate. In the game, an increase in heart rate compared to the initial baseline period, results in an increase in the computed arousal. To avoid that respiratory sinus-arrhythmia is reflected in the arousal value, the mean value over the last 5 heart beats is used for arousal computation (Berntson et al., 2007). For instance, when the heartbeat is 60 beats per minute, the time window for computing the arousal value is approximately 5 seconds. The arousal value is updated 60 times per minute (1 Hz).

Humans have limited information processing capabilities—particularly in fast-paced environments. Reducing the amount of information to a limited number of categories is thus essential to facilitate effective information processing (Miller, 1956). Previous research has shown particularly that reductions to five or seven categories are effective (e.g. Lozano et al., 2008). In order to provide the users with biofeedback in a meaningful way (cf. R3), the arousal parameter in the Auction Game can thus only take on five different integer values (1 to 5). A value of 5 corresponds to the highest level and is reached if the current heart rate is more than 15.00% higher than in the initial rest period. A value of 4 corresponds to the second-highest level and is reached if the increase in heart rate is between 11.25% and 15.00%. Finally, a value of 1 corresponds to an increase of 0.00% or lower. The threshold values were calibrated based on an iterative implementation process with student and real investor trial sessions. Moreover, the range of values is in accordance with the levels reported in previous research (Adam et al., 2013; Smith and Dickhaut, 2005).

7.4.3 Display of Arousal Values

The arousal parameter is displayed to the user by means of an arousal meter that visualizes the user's current emotional state. Meters are frequently used in user interface design to display important indicators in an intuitive and meaningful way (cf. R3, see (Ariely and Loewenstein, 2006; Caria et al., 2007) for a similar approach). The arousal meter visualizes the five different levels of arousal with colors ranging from green (1) to red (5). This approach is similar to the light patterns used in the Philips Rationalizer which were designed to make the arousal levels "intuitively clear" (Djajadiningrat et al., 2009, p. 44).

To objectively determine which game elements the users are paying attention to during the game, a Tobii T60 eye tracking device was used when demonstrating a prototype of the game to six students. The results revealed that users in general paid attention to the arousal meter. However, when the decision scenario became more complicated, the users hardly paid attention to the arousal meter. The users stated that this was due to lack of time during fast-paced decision–making. Based on these results, the visualization of arousal values was extended to also include the cloud estimations so that the important information is pre-

	Number of Participants		Physiological Measurement	Biofeedback	Influence on Game Difficulty	Age	Female
Evaluation Study I	36	[BF treatment: 19]	YES	YES	YES	23.39 [20-28]	12
		[NBF treatment: 17]	YES	NO	YES		
Evaluation Study II	68	[BF treatment: 44]	YES	YES	YES	22.06 [18-27]	16
		[NI treatment: 24]	YES	NO	NO		

Table 7.2: Overview on Evaluation Study I and Study II.

sented at the attention focus point. More specifically, the cloud estimates were implemented to change their color according to the current level of arousal. This feature was found to be very useful in subsequent demonstration sessions.

7.5 Evaluation

7.5.1 Experimental Design

To evaluate whether the Auction Game can induce high levels of arousal (R1) and whether effective Er is rewarded (R2), we conducted two laboratory experiments (Evaluation Study I and II). Moreover, by activating and deactivating biofeedback and the impact of the measured arousal levels on game difficulty, we evaluated whether the display of biofeedback was meaningful in the sense that it helped the users to perform better in the game (R3).

The core parameters of the two experiments are summarized in Table 7.2. Both studies are based on between-subjects designs with a total number of 36 and 68 subjects, respectively[4]. In the biofeedback treatment (Bf), the subjects are provided with a direct feedback on their emotional state. In the no-biofeedback treatment (Nbf), subjects do not gain any direct information on their emotional state. In both treatments, however, the arousal level has a direct impact on game performance (see design section). Thus, in both treatments the level of arousal is positively correlated with the game difficulty. The participants in Nbf and Bf receive the same instructions on how emotional arousal influences game difficulty, except that the Nbf version does not get any information on the arousal meter

[4]Originally, 46 subjects participated in Evaluation Study I and 80 subjects participated in Evaluation Study II. Because of difficulties with the data transmission, we had to exclude 10 subjects from Study I and 12 subjects from Study II.

and the adjusting color of the clouds, since these elements do not occur in their game version.

Finally, in the no influence (Ni) treatment, subjects play the game in a mode where neither the emotional state is indicated nor does it influence the game whatsoever. For this group also the instruction material differs slightly: Participants of the Ni treatment only received the information that "paying attention is an important task to regulate your emotions," without referring to arousal. It is important to highlight in this context that the arousal influencing the game can only increase game difficulty. This means that subjects in the Ni treatment played a game in which the trading task is never more difficult than it is in the Bf treatment. Therefore, it is interesting to analyze whether biofeedback and thus an increased sensitivity for Er has a positive effect on decision performance in the Auction Game.

7.5.2 Measures

Our evaluation is based on a set of specific measures. First, the users' current level of arousal is assessed on a scale from 1 to 5 by means of Ecg measurements as outlined in the Implementation section. Second, we measure each user's decision performance in the Auction Game. More precisely, a user's individual decision performance is defined as the percentage of decisions taken correctly out of all decisions. Pressing the no-trade button is defined here as a correct decision, since it always prevents against losing € 5.00 when not choosing one of the other options. Third, we assess to which extent Er strategies are consciously applied by individual users. Therefore, we use the Er questionnaire (Erq) by Gross and John (2003), which focuses specifically on the strategies cognitive reappraisal and suppression (see also (Abler and Kessler, 2009)). The Erq is characterized by high reliability, high discriminate validity, and only little correlation with the dimensions of other psychological questionnaires (Kannan and Kopalle, 2001). We conduct a factor analysis and validity checks on the measurement scale of the ERQ[5]. Both strategies assessed with the Erq have been proven to be beneficial for regulating the emotional state. Therefore, we use a single measure that equally reflects both strategies as an indicator for consciously applying Er strategies; in the following denoted as Er score. Fourth, we directly asked the subjects of Evaluation Study I which strategies they applied in order to down-regulate their emotional arousal and analyzed the effectiveness of these strategies.

[5] All items load properly toward the respective constructs and have only little mean crossloading of 0.096. Cronbach's alpha for cognitive reappraisal was 0.834 and 0.745, for suppression. Cumulative explained variance is 0.471. The results are comparable with the findings of Gross and John (2003).

7.5.3 Experimental Procedure

Both experiments were conducted at the Karlsruhe Institute of Technology, Karlsruhe, Germany. The experiments were approved by the respective authorities and follow their ethical principles, which include nondeception of participants, noninvasive measurements, performance-based payments, and anonymous data collection. The subjects were informed in the invitation letter that their heart rate would be assessed during the sessions. In every session, we invited 30 percent more participants than actually needed in order to prescreen participants upon arrival in the laboratory. Evaluation Study I was conducted in late 2011 and Evaluation Study II in early 2012. In both studies we controlled for a possible effect of time by distributing the treatments in similar proportions across the morning and afternoon sessions. The average room temperature was 25.47 °C (77.84 °F). The ORSEE software environment was used to recruit participants from a pool of university students (Greiner, 2004). In the first study a constant conversion of game euros earned into real cash was used as an incentive to participate. Payment, based on performance, was a constant conversion of game euros earned times 0.05 into real euros; plus a fixed participation fee of € 5. In the second study the game was played within a set of games (which will not be described here in detail), and the payment was randomly chosen according to the performance in one of the games. Thus, in both studies the subjects had a high incentive to perform their best in order to earn money.

The subjects were randomly seated in the laboratory and the use of the sensors was explained during the registration. The subjects received an envelope which contained the number of their work station and their anonymous identification number. To reduce distractions, each participant sat in his or her own cubicle with noise-cancelling headphones and was therefore completely separated from the other participants. After the instructions were read to the subjects they started playing a tutorial that walked them slowly through the game mechanisms. After the subjects completed the training mode, there was a five-minute rest period before proceeding with the game.

7.6 Evaluation Results

In this section we present the results of the two evaluation studies along with the three requirements introduced in the design section. Altogether, subjects took around 79 decisions on average before the game stopped, with 21 decisions being the fewest and 177 being the most. The no-trade button was used in around 10% of the decisions and subjects needed on average 1.15 seconds to take a decision once they had received the three price estimates. Of all the 104 subjects who participated in the two studies 18 managed to complete the game.

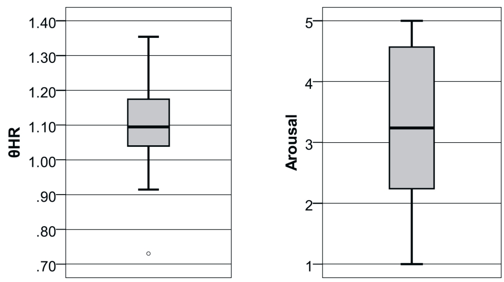

Figure 7.3: Boxplots of subjects' average heart rate and arousal. *Note:* The correlation between heart rate and arousal does not map perfectly, since θ HR values smaller than 1.00 result in arousal values of 1 and θ HR values exceeding 1.15 result in arousal values of 5.

7.6.1 Evaluation Results for Requirement R1

To examine the engaging impact of the Auction Game, the subjects' arousal level and heart rate during the game were constantly assessed. The arousal parameter reveals that the game in fact induces arousal: the average arousal level, which can take on values from 1 to 5, is 3.272 (SD=1.285). For most of the subjects (97%) the highest arousal level of 5 was reached at some point of the game, and for the majority of subjects (65%) all possible values from 1 to 5 were encountered during the course of the game. Figure 7.3 illustrates the subjects' average arousal and heart rate during the game relative to the baseline period (θHR). For the large majority of subjects (88 percent), the average heart rate is higher during the game period than it is in the baseline period. A paired-samples t–test shows that the subjects' heart rates during the game are, on average, 11 percent higher than in the baseline period (1.00 versus 1.107, n = 104, t(103) = –10.351, $p < 0.001$).

These increases in heart rate are remarkable with respect to the fact that the subjects did not engage in any physical exercise, but simply played an engaging serious game. This is also supported when comparing our results to other studies. Adam and colleagues also employed HR as a measure for physiological arousal in fast dynamic auctions, but report slightly lower values (Adam et al., 2013). Smith and Dickhaut list typical heart rate responses to induced idiosyncratic emotions (Smith and Dickhaut, 2005). The reported scores average for the

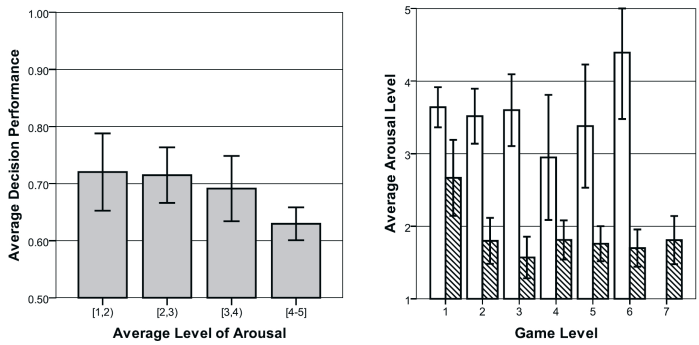

Figure 7.4: Average Decision Performance and Average Arousal Level. *Notes:* The left bar chart illustrates subjects' average decision performance depending on their state of arousal. The bar chart on the right indicates subjects' average arousal during the game levels (Error bars: 95% confidence interval).

strongest emotions around 4 to 7 beats per minute above baseline. This is lower than the average heart rate during the Auction Game, which results on average in eight beats per minute above baseline.

In Evaluation Study I, we asked the subjects to fill out a questionnaire and report the affective states they had experienced during the experiment (from 1 = "not at all" to 7 = "very much"). The highest reported emotional states were tension (M=4.78; SD=1.290), joy (M=4.08; SD=1.592), and anger (M=3.97; SD=1.828). Additionally, subjects were asked to express their overall impression of the game in a free text field. The majority of the subjects reported that they liked the game and also felt highly aroused while playing it. They liked to pick up the concept to regulate their emotions in order to ease game difficulty. These self-assessments are in line with our findings from physiology that the game is capable of inducing arousal. Based on the physiology and questionnaire results, we conclude that in line with R1 the Auction Game provides an engaging learning environment which can elicit arousal and, by design, is embedded in the context of financial decision–making.

7.6.2 Evaluation Results for Requirement R2

To test whether the game rewards effective Er, we analyze whether those subjects who managed to down-regulate their arousal are in fact more successful in the game. For the evaluation of R2, we therefore focus only on those subjects for

	Decision performance			
Independent variables	B	SE	t-statistic	Significance
Arousal	–0.033	0.010	–3.462	< 0.001***
ER score	0.033	0.016	2.123	0.037*

Notes: Regression for subjects' decision performance on arousal and ER (n = 80). * p < 0.05; ** p < 0.01; *** p < 0.001.

Table 7.3: Regression Results for Requirement 2

which arousal actually had an influence on game difficulty (n = 80). As depicted in the left-hand side of Figure 7.4, the subjects' average decision performance decreases with increasing levels of arousal. To test this relationship, we conduct an ordinary least squares (OLS) regression, which is summarized in Table 7.3. The regression confirms that arousal has a significant negative influence on decision performance (B = –0.033, SE [standard error] = 0.010, t = –3.462, p < 0.001). Correspondingly, the right part of Figure 7.4 shows that the average arousal level is lower for those subjects who managed to reach the final game level (n = 18) compared to the other subjects.

We assessed whether effective ER strategies were rewarded while playing the game by analyzing the results of the ER questionnaire, which we used to determine subjects' consciously employed tendency to regulate their emotions. As can be seen in Table 7.3, the regression reveals a positive effect of the ER score on decision performance (B = 0.033, SE = 0.016, t = 2.123, p < 0.05). Interestingly, subjects' score on ER is also related with their tendency to hit the no-trade button and thereby escape the € 5.00 penalty (n = 80, Pearson's r = 0.213, p = 0.058). However, this effect is only marginally significant.

Moreover, in Evaluation Study I we used a free-text questionnaire to ask the users what kind of strategies they used to actively down-regulate their arousal. Out of the 36 participants, 21 stated that they tried to regulate their breathing in order to regulate their arousal. Common answers were "I tried to breathe calmly" and "I concentrated on breathing." Two subjects indicated that when they realized they were too aroused, they closed their eyes for a moment in order to take a step back from active trading and relax before returning to it. Six subjects tried to "think of something different." These strategies are similar to professional meditation techniques such as applied in mindfulness interventions, which are often linked to ER (Kabat-zinn et al., 1992). Active breathing and closing eyes led to an increased average decision performance in the game. We also find that the subjects' experience of subjective "frustration, when making a wrong decision" positively correlates with arousal (n = 36, Pearson's r = 0.374, p <0.05).

7.6.3 Evaluation Results for Requirement R3

By design, biofeedback in the Auction Game is based on unobtrusive measurements (see the Design section). To test whether biofeedback is also provided in a meaningful way, that is, whether the users can actually draw useful information from it, we analyze the subjects' decision performance with and without biofeedback.

In Evaluation Study I, one group played the Auction Game with biofeedback (Bf) and one group played the Auction Game without biofeedback (Nbf). In other words, the subjects in the Nbf treatment did not see the arousal meter and color adjustments of the clouds. It is important to note, however, that the positive correlation of arousal and game difficulty remains unchanged, only the feature of biofeedback is controlled for. As a first descriptive result, it is interesting to note that subjects in the Bf treatment stated that they do not think that the biofeedback helps them either to "become aware of their current emotional state" (M = 2.65, SD = 1.656) to "help them with making better decisions" (M = 1.82, SD = 1.074). However, a t-test confirms that the subjects of the Bf treatment achieved a better decision performance than the subjects of the Nbf treatment (0.749 versus 0.658, n = 36, t(34) = 2.530, $p < 0.05$). Taken together, these results demonstrate that biofeedback helped the subjects increase their decision performance in the game.

In Evaluation Study II, we introduced a no-influence (Ni) treatment in which the biofeedback was neither displayed nor influenced game difficulty. In other words, the two core features of the game "live biofeedback" and "arousal influencing game difficulty" were deactivated in the Ni treatment. In this context, it is important to note that the correlation of arousal and game difficulty can only make the game harder to play. Therefore, in absolute terms, the subjects in the Ni treatment played a game in which the trading task is never more difficult than it is in the Bf treatment. However, if the subjects can employ the biofeedback profitably, this might mitigate or even offset this effect. The results of Evaluation Study II show that the subjects in the Bf treatment performed better in the Auction Game than those subjects who were in the Ni treatment (0.688 versus 0.629, n = 68, t(66) = 1.985, $p = 0.051$); however, this effect is only marginally significant. In other words, those subjects who knew that their arousal level influences game difficulty perform better in the game—even though playing a more difficult game. It therefore seems plausible that the positive effect of biofeedback offsets or even exceeds the negative effect of the correlation between arousal and game difficulty. It is important to note, however, that we cannot control for game difficulty here, as it is by design endogenously connected to arousal. Based on our study, we thus cannot draw general conclusions on how biofeedback affects decision performance in general. We will return to this point in the discussion. In summary, we conclude that the users were provided with biofeedback in a meaningful way in the sense that it helped them to perform better in the Auction Game.

This is also supported when comparing the overall arousal level in the two treatments. The overall arousal level was significantly lower in the BF treatment than it was in the NI treatment (2.871 versus 4.030, n = 68, t(66) = -3.820, $p < 0.001$). Thus, the subjects were able to draw information from the biofeedback and to down-regulate their current level of arousal. Based on the results of Evaluation Studies I and II, we conclude that in line with R3, the Auction Game provides the users with live biofeedback in a meaningful way, which is, by design, based on unobtrusive physiological measurements.

Requirement	Design Decision	Evaluation Result
R1: *The design artifact has to provide an engaging learning environment which can elicit arousal and is embedded in the context of financial decision making.*	- To provide an engaging learning environment, a serious game with different game levels was designed. Arousal is induced by limited time to take a decision, limited information display, limited time to finish the level, arousing music that speeds up with level progress and arousing pictures (taken from the IAPS database). - The game is embedded in a financial decision scenario. The user takes on the role of a trader and constantly conducts buy and sell decisions based on previous price information. - Game difficulty and arousal parameter were calibrated in demonstration sessions with practitioners and student testers.	- Based on the physiology and questionnaire results, we conclude that the game provides an engaging and arousing environment. The elicited physiological responses are similar to other studies that induce emotional stimuli. Therefore a necessary prerequisite to actively practice down-regulation of arousal is achieved. - The financial decision scenario is of limited complexity and therefore easy to pick up for users even in the first session. Practitioners from the field report that they find similarities to their daily tasks in the game.
R2: *The design artifact has to provide an environment in which effective emotion regulation is rewarded.*	- Game difficulty is directly linked to the player's current emotional state (operationalized as level of arousal). - Variance of price estimates changes depending on the current arousal level. - Heart rate is used for calculation of arousal, as the game is characterized by a fast-paced decision environment. - Users are given the opportunity to actively decide to take one step back from trading in order to calm down before returning to active trading (no-trade button).	- Increasing levels of arousal are negatively correlated with decision performance in the game. - The emotion regulation questionnaire (ERQ) is used to determine subjects' tendencies to apply emotion regulation strategies. Applying emotion regulation strategies correlates with game performance. The game therefore rewards the effective use of emotion regulation strategies. - Emotion regulation is rewarded and can be practiced in the game.
R3: *The design artifact has to provide the user with live biofeedback in an unobtrusive and meaningful way.*	- Biofeedback is based on unobtrusive ECG-measurements with dry electrodes and wireless data transmission. - Calculation of an arousal parameter that can only take on values between 1 and 5 to reduce information load. - Use of an arousal meter and colors to make the biofeedback more intuitive and thereby more meaningful.	- Those subjects who played the game with biofeedback managed to achieve on average a better down regulation of arousal and decision performance than those subjects who played the game without biofeedback. - Subjects perform better in the game when they are provided with a visualization of the biofeedback, although they state that this information has no influence.

Table 7.4: Summary of requirements, design decisions, and evaluation results.

7.7 Discussion and Conclusions

7.7.1 Summary of Results

In this paper, we designed, implemented, and evaluated a biofeedback based NeuroIs tool aimed at supporting decision makers with improving their emotion regulation capabilities—the Auction Game. To the best of our knowledge, this is the first serious game in which online biofeedback is directly applied to a financial decision–making context. In addition to the system of Philips and Abn Amro, which displays biofeedback by means of a light bowl next to the monitor, our design artifact integrates biofeedback into the decision scenario and thereby aims at improving the users' emotion–regulation capabilities. A core feature of our approach is that game difficulty is directly linked to the user's current emotional state. The emotional state is assessed by means of unobtrusive Ecg measurements. Depending on the player's ability to regulate emotions, the financial decision scenario of the game continuously adjusts itself and thereby becomes more (or less) difficult. In order to evaluate the game and provide further insight into Er processes, we conducted two evaluations studies. Table 7.4 summarizes the evaluation results along with the requirements and design decisions.

The results of the evaluation studies show that the game in fact induces arousal and rewards down-regulation of arousal. Moreover, it becomes evident that the game rewards the effective use of Er strategies. In summary, the Auction Game provides a learning environment in which Er can be practiced and rewarded. In addition to that, it was found that the biofeedback displayed in the Auction Game is meaningful for the users in the sense that it helps them to down regulate arousal and perform better in the game.

7.7.2 Managerial Implications

The Auction Game was designed and implemented in close correspondence with practitioners from the banking and finance industry. Therefore, it is particularly relevant for professional traders and private investors who want to improve their emotion–regulation capabilities. In this regard, the Auction Game has already drawn attention at several international trading and investment conferences. Especially during stressful periods traders and investors are highly motivated to remain level-headed and improve their emotion–regulation capabilities (Fenton-O'Creevy et al., 2011). The Auction Game provides market participants with a tool that can help them to train stressful periods and actively practice Er. Banks like Abn Amro, Barclays, and Saxo Bank have recently become aware of their customers' interest in de-biasing and ER training. In this context, a growing number of banks strive to offer their customers specific training services for skill development. The Auction Game and similar biofeedback approaches can thus be included as intervention elements. The Philips Rationalizer was an important first step in this direction (Djajadiningrat et al., 2009). Similar to the Auction Game, future trading systems will also directly incorporate biofeedback into their trading interfaces. But also beyond the scope of skill development and training, biofeedback will play a more and more important role for the designers of trading interfaces.

In the near future, shirts with built-in dry electrodes will come onto the market. Technical progress already allows wireless recording and real-time processing of ECG data. Soon it will be possible to constantly measure and process a wide range of physiological parameters in real-world decision scenarios instantaneously and integrate them into the trading interfaces, for instance, at day-trading centers and also for private investors at home. Thus, it will be possible to provide the decision makers with a highly sophisticated live biofeedback on their current emotional state, to immediately warn them in critical situations, and to support them with improving their emotion–regulation capabilities. In this context, it will be very interesting for traders, investors, and consumers to match arousal to their decision performance on a longer time base (e.g., several weeks or months). The continuous assessment of the emotional state during periods of strong and poor decision performance could help subjects to understand how their own arousal and decision performance are interlinked. Biofeedback games such as the Auction Game can then be used to help subjects to regulate their arousal to an individually optimal level—also in other contexts than financial decision–making.

While there is a growing number of related studies in the fields of Neuroergonomics and affective computing, the use of tools from cognitive neuroscience in design science research has only just begun (vom Brocke et al., 2012, p. 2). By applying methodologies from neuropsychology and psychophysiology to information systems research, the new field of NeuroIS provides long overdue insight into the decision–making process of human individuals when interacting with modern information technology (Dimoka and Banker, 2012; Dimoka et al., 2011). Thereby, one major goal is to support human decision makers. In this sense, the Auction Game is a use case for how information systems can incorporate physiological and neurological biofeedback in order to improve human decision–making. Applying ER strategies in the context of information systems research is not only limited to decision–making alone. It can also be helpful for well-being, learning, and health. Riedl and colleagues showed that IT users are exposed to "Technostress" (Riedl et al., 2012). Down-regulating arousal can help subjects to reduce the impact of stress and, thus, to maintain and improve health (Gross, 2009). Therefore, ER training and the integration of biofeedback into information systems are also highly relevant from a management perspective as it may help managers to reduce their employees overall stress levels and thereby increase productivity.

7.7.3 Limitations and Future Research

This study has several limitations. First, the Auction Game was designed to reward down-regulation of arousal. While it was shown that in many scenarios, arousal can impair decision–making (Kuhnen and Knutson, 2005; Peterson, 2007), there are also scenarios where it may be beneficial to up-regulate arousal to an optimal level and maintain that level (e.g., flow theory, Yerkes–Dodson law). In this respect, it is important to note that we do not pursue a normative approach in the sense that low arousal is always preferable. We rather aim at improving the user's capacity to reduce high levels of arousal when he or she aims at doing so. For future research, the design artifact could easily be adjusted with the goal to train up-regulating arousal or even to target an optimal level of arousal. Subjects' trading data from the field could then be assessed in order to detect an individual optimal arousal level, where performance is highest. Then, similar to how athletes warm up before sporting events, traders could warm up before they start with their actual task in order to reach their optimal arousal level.

A second limitation of this study is that because of the fast-paced decision environment and further reasons outlined in the Design section, the current calculation of arousal is based on heart rate measurements solely. Dependent of the particular context and focus, future research should employ further physiological parameters, such as skin conductance or electroencephalogram (EEG). Both of these measures are well-known physiological proxies for emotional processing. Also, noncontact monitoring of arousal has been successfully employed in other studies (e.g., (Nunamaker et al., 2011)). A combination of several parameters will produce a more accurate and robust computation of arousal.

A third limitation arises from the fact that we varied two features between the treatment groups in Evaluation Study II, the connection of game difficulty to arousal and the display of biofeedback. In both of our studies we found biofeedback to be meaningful in the sense that it helped users to perform better in the game. It is important to highlight, though, that we cannot directly draw inferences with respect to whether immediate biofeedback also generally increases decision performance. By design, arousal and game difficulty are combined in our approach. The question of whether biofeedback is beneficial for users beyond the use case of Er training is, however, highly relevant for Is design science research. A fourth limitation is that we did not take into account further measures for personality, such as cognitive ability or the Big Five personality traits (Gross and John, 2003; Sokol-Hessner et al., 2009). We believe that disentangling these effects is an important area of future research.

For future Is design science research it will be interesting to investigate to what extent biofeedback environments can (1) be of valuable decision support for the users and (2) contribute to a long term skill development of effective emotion. Also, other Er approaches, known to increase interoception—such as different forms of relaxation exercises, yoga, breathing patterns, or mindfulness courses—could then be evaluated with respect to their beneficial influence. In general, more biofeedback studies need to be conducted to help unravel its influence on decision–making. Are traders less susceptible to decision biases when they are provided with live biofeedback? Can live biofeedback help managers to maintain control of highly emotional situations? Do consumers behave less impulsive during online shopping when they are provided with live biofeedback? We believe that answering these questions can contribute to more sophisticated NeuroIs tools that support users in taking more advantageous decisions.

7.8 Conclusions

Taken as a whole, this study demonstrates the potential of biofeedback-based NeuroIS tools and derives a set of design guidelines for integrating biofeedback in IT artifacts. Generally speaking, emotions are an integral part of our lives that guide many facets of human experience. Who would want to live without emotions? However, although emotions are essential for making advantageous decisions, they can also get out of control. NeuroIS tools can help us to increase emotional awareness, to improve our emotion–regulation capabilities, and, in consequence, to avoid disadvantageous decisions with undesired consequences. In this sense, we understand this paper also as a motivation for Is researchers to incorporate the concept of Er in Is research.

7.9 Acknowledgments

The authors thank the Editor as well as the reviewers for their valuable feedback on an earlier version of the paper. They thank Christian Caspary, Eike Holst, Judith Mang, Xiaoyun Xu, and Christoph Zimmermann for their help with preparing and conducting the experiments, as well as Stefan Hey, the research

Figure 7.5: Game Screen of the Auction Game.

group hiper.campus, and the company movisens GmbH for their technical support. Financial support by the German Research Foundation Graduate School 895 "Information Management and Market Engineering" at the Karlsruhe Institute of Technology is gratefully acknowledged. This research was conducted within the project xDelia, funded by the European Commission under the 7th Framework Programme, Grant no. 231830. They thank the project colleagues, the project reviewers, and Meike Decker for their constructive comments on the design and development process. Financial support by the European Commission is gratefully acknowledged.

7.10 Appendix

The Auction Game was developed using the Unity 3D Pro game engine. The game is played in a 2D environment, where price estimates are presented inside colored cloud drawings. To depict a sense of progress through the game, every level has a different background picture of the sky.

Figure 7.5 displays a screenshot of the Auction Game. On the top–left side of the screen, the remaining time for the current level is displayed. The player can see his or her individual arousal level indicated on the meter in the right top corner, as well as by the color of the clouds (green, yellow, and red). The profit goal and total money earned are presented on the meter at the bottom-right side of the screen. Decisions can be made by clicking on the buy, the sell, or the no-trade button at the bottom of the screen. Note that in contrast to the illustration in Figure 7.5, the clouds appear sequentially in the game and that the location of their appearance on the screen is random. The decision terminal does not get activated before the last cloud got indicated. Every profitable decision will reward the player with a certain amount of money, while a nonprofitable decision will reduce the player's earning and take him or her further away from the current level's profit goal.

Before the game starts, the instructions on how to play the game are displayed. Moreover, the player can choose to go through a tutorial. The tutorial explains the principle of the game and introduces its different elements step by step. The tutorial should be played the first time the player gets in contact with the game, but it can be skipped if the player already knows the game. Starting from level 6, affective pictures from the IAPS database are randomly displayed to distract the user and further induce arousal. Thereby, we selected particularly pictures with high arousal scores while excluding disgusting pictures.

Because of the flexible software architecture of the xAffect framework, the Auction Game can be connected to a variety of measurement devices. In the evaluation studies described in this paper, we used two different types of measurement devices. The first device is called ekgMove and is produced by the company Movisens. It is a sensor on a chest strap with dry electrodes. The strap is attached directly on the subject's chest (comparable to the ones used during sport exercises). The assessment is performed using Bluetooth, and therefore the chest strap and sensor are very comfortable to wear as there are no wires required. According to the manufacturer, the time lag due to wireless transmission is approximately 50 milliseconds. The second device is called Varioport–e and is produced by the company Becker Meditec. In contrast to the ekgMove, it requires a constant cable connection of the electrodes to the PC. The Varioport-e system was used for measuring the ECG of eight subjects in Evaluation Study I.

Chapter 8

Modeling Cognitive Load and Physiological Arousal Through Pupil Diameter and Heart Rate

Abstract

This study investigates individuals' cognitive load processing abilities while engaged on a decision–making task in serious games, to explore how a substantial cognitive load dominates over the physiological arousal effect on pupil diameter. A serious game was presented to the participants, which displayed the on–line biofeedback based on physiological measurements of arousal. In such dynamic decision–making environment, the pupil diameter was analyzed in relation to the heart rate, to evaluate if the former could be a useful measure of cognitive abilities of individuals. As pupil might reflect both cognitive activity and physiological arousal, the pupillary response will show an arousal effect only when the cognitive demands of the situation are minimal. Evidence shows that in a situation where a substantial level of cognitive activity is required, only that activity will be observable on the pupil diameter, dominating over the physiological arousal effect indicated by the pupillary response. It is suggested that it might be possible to design serious games tailored to the cognitive abilities of an individual player, using the proposed physiological measurements to observe the moment when such dominance occurs.

8.1 Introduction

Research shows that emotions impair or facilitate decision–making (Hu et al., 2015; Adam et al., 2015). Russell's model (Posner et al., 2005) classifies emotions through valence and arousal, their independent components. Therefore, valence represents whether the current emotional experience of a situation is positive or negative. On another hand, the level of excitement is represented by arousal. Following this classification, emotions are measured through the valence and arousal components. The concept of physiological arousal has been generally validated in models of emotion (Porges, 2007). Task difficulty is correlated with physiological arousal (Cohen, 2011), thus a challenging decision task might be reflected in physiological signals. The author further states that there is a moment when performance decreases as arousal continues to increase. The pupil dilation is influenced by physiological arousal, together with attention and interest (Benikos et al., 2013; Andreassi, 2013). The pupil diameter (PD) has been validated as a useful measure of physiological arousal (Laeng et al., 2012; Granholm and Steinhauer, 2004). Early investigators (Hess and Polt, 1964) viewed the pupillary response as reflecting the level of arousal or emotionality; followed a few years later (Hess and Polt, 1964; Kahneman and Beatty, 1966) by studies arguing that the pupillary response might reflect cognitive activity as well as or instead of arousal. Thus, Stanners et al. (1979) concludes that the pupil response will show an arousal effect only when the cognitive demands of the situation are minimal. If the situation requires a substantial level of cognitive activity, only that activity will be observable on the pupil diameter, dominating over the physiological arousal effect indicated by the pupillary response (Pehlivanoglu et al., 2014). These findings give rise to a need for the investigation of the moment when the dominance occurs, due to substantial cognitive demands of the task. Moreover, it indicates a need for deeper understanding of how cognitive load affects the physiological arousal effect, in an attempt to inform the design of such tasks. Furthermore, just like in the effects observed in the pupil, an increased task difficulty increases the heart rate (HR) (Boutcher and Boutcher, 2006; Sosnowski et al., 2004). So the same suggestion for the pupil is applicable to the heart, where it is unclear to which extent is the cognitive activity uniquely associated with the physiological arousal response (Steinhauer et al., 2004).

This investigation is motivated by these findings, as it follows up on the previous study (Steinhauer et al., 2004) to include also an on–line biofeedback presentation, where participants have been engaged in a decision–making task. The identification of cognitive load through the pupil is already recognized in the community (Hung et al., 2017). The authors argue that traditional physiological methods of detecting arousal need to be enhanced with cognitive load of individuals, as the cognitive load can affect physiological changes and mask the arousal information in data. Moreover, it can have a negative effect on players' emotions, which might disrupt the acquisition of the intended skill. Recent findings in the serious game design take into consideration an individual participant information to provide an enhanced experience (Charles and Black, 2004), moving towards player–centered design (Charles et al., 2005). Due to the differences in personality and skill, diverse players would have different performance on the task in serious games. For some players whom the challenge is too high, this would either reduce performance on the task, push them towards an unwanted emotional state, or disrupt the acquisition of the skill they are trying to master. Depending on the design goals of serious games, this might be unacceptable. Therefore, it may be beneficial to provide a feedback for the participants to shape awareness of their own cognitive abilities and to tailor serious games to the cognitive abilities of an individual player. Usage of unobtrusive on–line physiological signals is necessary to asses the substantial cognitive load of players engaged on a decision–making task in serious games. Since pupillary response will show an arousal effect only when the cognitive demands of a situation are minimal, these modalities may provide the information of the moment when the dominance of the cognitive load over the physiological arousal effect occurs. Players might be engaged in a challenging situation at such moment, where game designers might take action to mitigate the disruptive effects of such states and change the difficulty of a task for each individual player to achieve the design goals of serious games. Therefore, the research question this study focuses on is to find the moment when substantial cognitive load overshadows the arousal effect, in an attempt to investigate cognitive abilities of the participants engaged on a decision–making task. Following those propositions, an infrared eye–tracking system and an electrocardiograph (Ecg) sensor emerged as a solution providing quantitative measurements of the moment when substantial cognitive load overshadows the arousal effect (Steinhauer et al., 2004). Therefore, a dynamic serious game environment with an on–line biofeedback has been used. These insights could provide a deeper understanding of how physiological arousal is underlying cognitive load in such decision–making tasks from the human participant perspective, potentially informing the design of more meaningful serious game tasks that would be tailored to the cognitive abilities of an individual interacting with them.

The remainder of the paper is structured as follows: background is given in Section 2, Section 3 presents the experimental set–up and methodological approach. Results are given in Section 4, discussion and limitations are given in Section 5 and Section 6 respectively. Conclusions are detailed in Section 7.

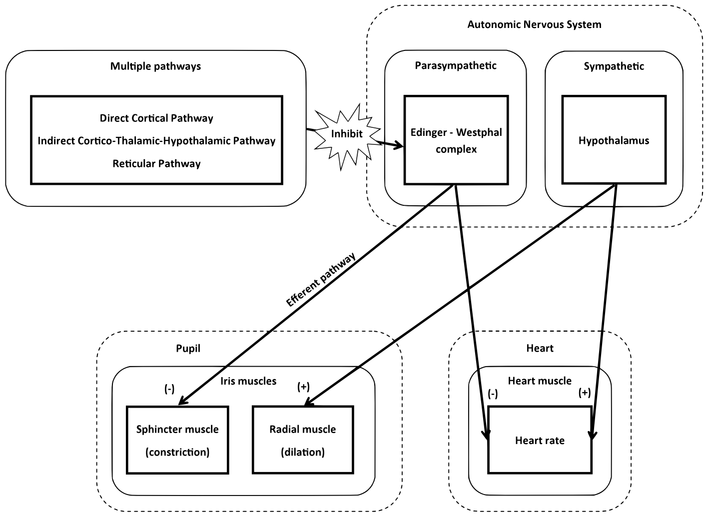

Figure 8.1: Model of the ANS control pathways for the pupil diameter and the Hr.

8.2 Background

Serious games can be defined as games with a purpose other than just entertainment, where they are aiming at captivating and engaging a participant for a specific purpose (Léger et al., 2011; Susi and Johannesson, 2007). They are rewarding investigation platforms since participants must explore their options, even though complete or clear information about the best course of action is not available (Léger et al., 2011). Serious games support higher motivation on the task, as it is significantly correlated with the players' ability to handle cognitive load (Su, 2016). The authors state that the individuals with higher motivation can handle a task using lower cognitive load processing ability. Physiology integrates continuous and unobtrusive measures of physiological arousal in serious games (Nacke et al., 2009). Such physiology metrics offer insights into the participant's experience (McAllister et al., 2013), and behavior (Guardini and Maninetti, 2013; Sennersten and Lindley, 2009). Among a variety of eye–tracking techniques, pupil diameter is already a familiar metric for the experience and behavior in serious games, due to its ecological validity and ease of use (Lindley et al., 2008).

8.2.1 Eye Pupil and Responses

The pupil is an opening at the center of the iris surrounded by the iris muscles that respond to outside stimuli to control the amount of light that reaches the retina. We differentiate between two types of iris muscles: the sphincter muscle *sphincter pupillae* forming a band around the inner margin of the pupil and radial muscle *dilator pupillae*, see Figure 8.1. Two synergistic pathways control the pupil size and dynamics. Those pathways are the parasympathetic pathway and the sympathetic pathway, operating on the smooth muscles of the pupil. The parasympathetic pathway is mediated with the efferent pathway originating in the *Edinger–Westphal oculomotor complex* in the midbrain and innervates the sphincter, which is the circular muscle responsible for the constriction, as seen in the reflex reaction to light. Inhibition of *Edinger–Westphal complex results* in relaxation of the sphincter muscles and, thus, significant dilation. The sympathetic pathway, mediated by the hypothalamus, innervates the radial dilator muscle of the iris responsible for dilation (Steinhauer et al., 2004).

Thus, Stanners et al. (1979) concludes that the pupil response will show an arousal effect only when the cognitive demands of the situation are minimal. If the situation requires a substantial level of cognitive activity, only that activity will be observable on the pupil diameter, dominating over the physiological arousal effect indicated by the pupillary response. In short, minimal cognitive load results in the physiological arousal effect present in the pupil response. Substantial cognitive load results in the pupil response overshadowing the physiological arousal effect. Steinhauer et al. (2004) gave neuroscientific support for this theory where they report that multiple pathways impinge on the *Edinger–Westphal complex*, resulting in pupillary dilation through inhibition of the parasympathetic pathway. Those multiple pathways that generally increase in inhibition across all task conditions may well include contributions of both direct cortical/indirect cortico–thalamic–hypothalamic pathways (Lowenstein, 1955) and reticular pathways contributing to arousal (Steinhauer et al., 2004). It has been shown that the pupil size reflects processing load or mental effort (Moresi et al., 2008; Recarte and Nunes, 2003). Thus, Steinhauer et al. (2004) state that demanding cognitive load, most likely associated with frontal cortical functioning, contributes heavily to this inhibitory process dominating over the arousal effect.

Granholm and Steinhauer (2004) state that the changes in pupillary motility have been noted and employed as indicators of emotional arousal. A proposition came from Partala and Surakka (2003) to use pupil size variation as a computer input signal in affective computing, for example. Bradley et al. (2008) present a strong case supporting that the pupil's response during affective picture viewing reflects emotional arousal associated with the increased sympathetic activity.

They argue that emotional arousal is a key element in modulating the pupil's response. Furthermore, they state that rather than varying with cardiac deceleration mediated by the differences in parasympathetic activity, they have found that pupillary changes during emotionally arousing picture viewing predominantly reflects sympathetic nervous system activity. While in the studies exploring effects of mental load and sustained cognitive processing on pupil change reported in the previous paragraph, the observed pupillary dilation appears to be mediated by parasympathetic inhibition of the sphincter muscle. These findings support the hypothesis that pupillary changes during the affective picture viewing are mediated by increased sympathetic activity and strongly suggest that emotional arousal affects the pupil diameter, independent of whether the pictures are pleasant or unpleasant in valence. These insights support the suggestion (Steinhauer et al., 2004) that it is unclear to which extent is the cognitive activity uniquely associated with sympathetic or parasympathetic nervous system activity. The authors further argue that, if we are to provide quantitative measurements of activation underlying the central nervous system mechanisms, it is important to differentiate between activation of these pathways. With that knowledge, one can evaluate the neurophysiological systems that contribute to cognitive activities by monitoring pupillary dynamics.

8.2.2 Heart and Physiological Arousal

Berntson et al. (2007) state that the Hr is a measure of both the sympathetic and parasympathetic autonomous nervous system activity. Correspondingly, sympathetic activity tends to increase Hr, while parasympathetic activity decreases it. This suggests that this sympathetically mediated response covaries with emotional arousal (Lang et al., 1993). In contrast, cardiac deceleration is generally greater when viewing unpleasant, compared to either pleasant or neutral pictures (Bradley et al., 2001). This suggests a parasympathetic activity (Berntson et al., 2007), as shown in the pharmacological blockade studies of fear bradycardia in animals. The term bradycardia defined as Hr which falls below 60 beats per minute (Bpm). Similarly to the pupil, Berntson et al. (2007) further report that the extent of interplay between sympathetic or parasympathetic nervous system activity provides a measure of a subject's emotion–regulation capabilities. Studies have found a strong correlation between the Hr and physiological arousal (Andreassi, 2013; Xu et al., 2010). Just like the effects observed in the pupil, it has been found that increased task difficulty (e.g., difficulty of a mental arithmetic task) increases the Hr (Boutcher and Boutcher, 2006; Sosnowski et al., 2004). Ecg has been successfully used as a measure for assessing physiological emotions in humans with more than 80% success rate (Al Osman et al., 2014;

Rani et al., 2007). The method of biofeedback using HR became ubiquitous for targeting towards enhancing individuals' abilities (Al Osman et al., 2014).

This investigation presents a dynamic serious game environment with an on–line biofeedback, where participants have been engaged in a decision–making task. Using such serious game together with an infrared eye–tracking system and an ECG sensor allowed for the assessment of physiological arousal and underlying the cognitive load in players engaged on the decision–making task. Therefore, the research question this study focuses on is to find the moment when substantial cognitive load overshadows the arousal effect, in an attempt to investigate cognitive abilities of the participants engaged on a decision–making task in serious games. These insights could provide a deeper understanding of how physiological arousal is underlying cognitive load in such decision–making tasks, providing a feedback for the players to shape awareness of their own cognitive abilities and tailor serious games to it.

8.3 Methodology

A regression relationship study was conducted where the participants interacted with a serious game while the continuous measurements of the HR and the PD were made.

8.3.1 Participants

Twenty-one students participated in the experiment. The age range of the participants was between 20–24 years. Participants have not reported any ophthalmological problems (other than corrected vision), also no psychiatric or major medical disorders. Participants were given a movie ticket as a reward for participating. All participants received complete information on the study's goal, experimental conditions and gave their informed consent. The experiments were carried out by the Game Systems and Interaction Laboratory (GSIL) at Blekinge Institute of Technology, Karlskrona, Sweden. The experiments in this study were conducted under the advice of The Local Ethical Committee, while the ethical approval was obtained through the xDelia project.

8.3.2 Experimental Setup

Participants were seated in a recliner chair in a small, sound–attenuated room. Lighting and temperature were controlled in such a way that artificial fixture light

Figure 8.2: The Auction Game – a serious game used in this study presents a challenging decision–making task to a participant, directly linked to his/her physiological arousal state indicated via the biofeedback (Jerčić et al., 2012). The biofeedback information regarding physiological arousal was given through the meter in the upper right corner and through the color of the clouds.

was used throughout the experiment, while the temperature was held constant at 23 °C ± 1 °C. The chest band holding a physiological *Movisens ekgMove* ECG sensor was attached. Participants were seated in front of and facing the *Tobii T60* eye–tracker at a 50–60 cm distance from the screen.

Physiological signals were acquired using the *Tobii T60 Eye–tracker* for the PD data, with data acquisition frequency of 60 Hz. In addition, ECG was acquired at the chest using the *Movisens ekgMove* sensor, which records signals with high accuracy, while still providing a less obtrusive measurement to most other commercially available ECG devices by using flexible belt with dry electrodes, with two contact electrodes between the left and the right lowest rib points. The raw ECG signal is transmitted via Bluetooth to the algorithm module processing the current arousal level of the players from the HR information (Schaaff et al., 2011). The arousal levels are computed in relation to a baseline period recorded before the game starts, and transferred to the Auction Game. The PD was measured using an eye–tracker, whose data were analyzed offline. Computation of the HR from the ECG data provided information about the current physiological arousal level of the participants. Those arousal levels were computed against the baseline period, which was recorded for five minutes in a resting state before the game starts.

Following the definition of serious games as games with a clear purpose other than just entertainment (Susi and Johannesson, 2007), The Auction Game (Jerčić et al., 2012) was used as the serious game in this study to investigate cognitive abilities of the participants and the moment when a substantial cognitive load overshadows the physiological arousal effect. The Auction Game has emerged as an investigation platform since it presents a challenging decision–making task to the participants, interconnected with the ECG sensor for the on–line assessment of physiological arousal and the biofeedback presentation based on HR values. An advantage of such a coupling between the Auction Game and the physiological ECG sensor was that it measured reliable physiological arousal in a stressful environment (Astor et al., 2013b; Jerčić et al., 2012; Schaaff et al., 2011). The physiological arousal continuously adjusted the decision–making task and thereby making it more (or less) difficult, depending on the participant's ability to regulate physiological arousal. The participants were given information regarding their physiological arousal through the meter in the upper right corner and through the color of the clouds, see Figure 8.2. Furthermore, they were informed that regulating their arousal would benefit them on the task presented.

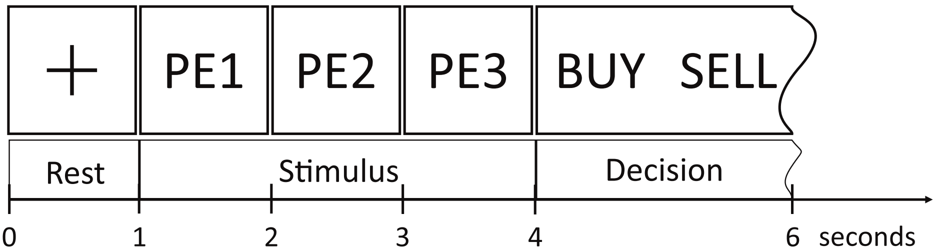

Figure 8.3: A trial consisted of three parts: (1) a rest period, where the participant is allowed to rest before the stimulus onset (0 – 1 s); (2) a stimulus onset period, where the participant is shown the three price estimations from which a correct price should be calculated that they should use to make a decision (1 – 4 s); and (3) an decision period, where the participant is making a buy/sell decision–based on the correct price and the offer given (4 – 6 s). The length of the decision period was dependent on the participant's level of arousal and getting even shorter as the trials progressed, minimum being one second.

8.3.3 Experimental Task

The participants were presented with a decision–making task to calculate a mean value from three given price estimations, to be able to reach a buy or sell decision at the correct price, see Figure 8.3. To make a decision, a participant had to click on the buy or sell button on the screen. Price estimations were directly linked to the physiological arousal level, such that they deviated from the correct price with higher variance the more aroused the participant was. Thus, a lower physiological arousal would make the variance of price estimations closer to the correct price, so that a buy or sell decision could have been easier. The serious game was linear, which meant that there was always just one possible correct decision to be made in each trial, and it was a reasonable challenge for most of the participants. Good emotion–regulation skills would benefit the participants to make the decision–making task less challenging, as higher physiological arousal also reduced the decision time, to promote a greater challenge. Moreover, the task became more challenging at subsequent trials as the decision period was further reduced, forcing a quick decision. On average, subjects performed 71 trials before the game stopped, with 19 trials being the fewest and 160 being the most.

The decision–making task was presented in the following steps, shown in Figure 8.3. Prior to the onset of price estimations, the participants had one second pause. Every price–estimation signal was displayed on the screen for one second, as a rule of thumb in fast–paced decision–making. The participants then had two seconds to reach a buy or sell decision, and shorter as the trials progressed or physiological arousal increased. After a decision, the feedback conveying the outcome of a decision was presented. The participant traded fictional 'goods' and only the correct price was relevant for the participant. For each traded good the participants received three price estimations indicated on the screen. The mean value of these price estimations was the correct price for reaching a buy or sell decision for an offered price. Depending on a participant's decision, s/he could have earned or lost virtual money (on average 2.50 €) relating to a correct or incorrect decision. The decision had to be made within the allocated amount of time, otherwise a participant was punished with a loss (-5 €). Such player–centered serious game design using biofeedback where physiological arousal was linked to the time and difficulty of the decision–making task allowed each individual participant to regulate their emotions and perform their maximum on the decision–making task reducing the influence of arousal. This allowed the observation of the participants' performance on the decision–making task limited by his ability to handle the cognitive load, reducing the influence of physiological arousal.

8.3.4 Protocol

Upon arrival, the participants were given general information about the experiment and the description explaining the serious game. Their written consent was obtained, explaining physiological measurements and data confidentiality. When the participants agreed to take part in the study, they signed an informed consent form. Each experimental session took around 20 minutes to complete.

1. Prior to the experimental session, a calibration of the eye–tracker was performed where the participants had to focus attentively on four individual focus points distributed randomly on the screen, after which the participants were instructed to start with the experiment. The luminance of the screen and the light conditions around the screen were measured and controlled for all trials.

2. Eye-tracking baseline recording of the individual PD variations was obtained from the three still color images (white, gray and black) presented for 60 seconds each, counterbalanced between the participants in order to minimize the ordering effect.

3. Practice run of the serious game was rehearsed in order to acquaint the participants with the task.

4. Arousal baseline for the HR measurements was taken in the resting phase prior to the game's onset while the fixation cross was presented in the middle of the screen. Participants were directed to relax for five minutes while the HR baseline recording was acquired.
5. The serious game was executed as described in the Experimental Setup section, presented in the Figure 8.3.
6. Following the experiment, the ECG sensor was removed. The experimenter subsequently debriefed and thanked the participant. The participants were given a movie ticket for their participation.

8.3.5 Data Analysis

Offline analysis was performed to obtain the PD values in relation to the HR values, to quantify physiological arousal data. In the offline analysis, HR data were used purely as a measure of the task difficulty.

Heart rate

The raw ECG signal was amplified, filtered using a band–pass filter at 10–40 Hz, and 16–bit digitized, the signal was then smoothed by a 10 ms moving average window. The R–peak (heartbeat) detection was based on OSEA algorithm (Hamilton, 2002) and it consisted of identification of R–waves applied to the ECG signal to obtain the interbeat intervals which were then reduced to HR, measured in BPM. The peaks were then detected in the resulting signal and detection heuristic rules were applied to avoid missing R–peaks or detecting multiple peaks for a single heartbeat. Artifacts were removed following the recommendations by (Clifford, 2007), where a minimum interbeat interval of 300 ms and maximum interval of 1500 ms between peaks were enforced, constraining the interval to be 20% different from the previous one. Furthermore, the peaks were checked for both positive and negative slopes to ensure that the baseline drift is not misclassified as a peak, while the detected missing peaks were backtracked with reexamination/interpolation. Such interbeat intervals were analyzed for ectopic beats (disturbance in the cardiac rhythm), arrhythmic events, missing data and noise effects outliers, which were linearly interpolated.

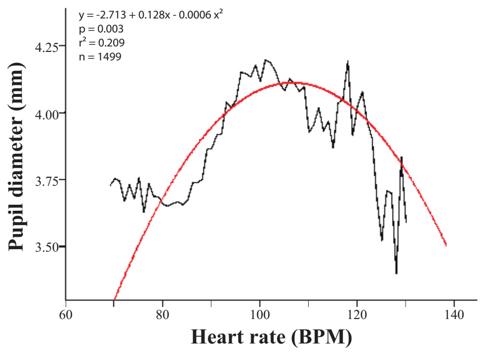

Figure 8.4: The normalized pupil diameter responses in relation to the HR on a decision–making task in a serious game. The substantial cognitive load dominating over the physiological arousal effect is observable on the PD as the HR increases. The black line represents the averaged PD values in relation to HR across participants and trials in the experiment. Red line represents best–fit non–linear quadratic regression curve for the PD values in relation to HR for all participants and trials over the whole experiment with 95 % confidence interval at the .003 probability level.

Pupil diameter

Regarding pupil data, off–line analysis of individual trials was carried out for the PD values in millimeters. The data were corrected for the short and long blink periods. A linear interpolation was then applied to the short blink periods. Baseline diameter was defined as the average diameter during the 60 s still image presentation (i.e., white, gray, and black screens), which was used to normalize the pupil diameter data used in the subsequent PD analyses. The PD maxima, minima and the mean absolute values for each eye were averaged across all the participants for the baseline measurements of each pupil and each participant, indicating where the center values are. The order of the presentations was counterbalanced between participants in order to minimize ordering effects.

Normalized PD values were grouped according to the accompanying HR values recorded. The normalized PD values were then averaged over HR values and plotted as the graph, shown on Figure 8.4. The PD values were normalized for each participant using the $BL_{min} = 1.34mm$ and $BL_{max} = 6.15mm$ obtained from the baseline period, using the normalizing Eq. (1).

$$PD_{norm} = \left(\frac{DP_t - DP_{min}}{DP_{max} - DP_{min}} \right) \times (BL_{max} - BL_{min}) + BL_{min} \qquad (8.1)$$

8.4 Results

Cognitive load and physiological arousal on a decision–making task were investigated to find the moment when one prevails over the other. To perform this investigation, the PD data were analyzed in relation to the HR data during this challenging task. A statistical post–hoc power analysis was performed for that relationship using the sample data from the experiment. The effect size in this study was ($N = 1499, f^2 = .264$), considered to be medium using Cohen's (Cohen, 1988) criteria. With an alpha level of $p < .05$ the statistical power for this study exceeded .99 for detecting a medium effect. Data were analyzed for non–linear regression in the SPSS statistical software package to estimate the relationship between these data variables. The data showed no violation of normality or homoscedasticity (data were normalized and the artifacts were removed before the analysis).

Data showed the maximum PD occurring across the HR values for the decision–making task. This coupling allowed for the observation of a substantial cognitive load dominating over the arousal effect on PD, and their dominance over each other, as determined by a non–linear regression on the values of PD data in relation to the HR data for the overall number (n = 1499) of measurements in the whole experiment. The data had a quadratic component that can be described by a best–fit curve having a non–zero slope. The regression analysis calculated the equation of the best–fit curve as $y = -2.713 + 0.128x - 0.0006x^2$ with $max(y) = 4.11mm$ occuring at $x = 106.67$ BPM ($n = 1499, p = .003, R^2 = .209$). The best–fit curve showed the moment when the dominance occurs with the PD maximum of 4.11 mm occurring for the HR of 106.67 BPM. It has been overlaid over the data averages across all participants and trials, shown on Figure 8.4 in red and black respectively. Moreover, the Figure 8.4 shows an interesting relationship between PD to HR, which greatly resembles the Yerkes–Dodson reversed 'U' curve (Cohen, 2011).

8.5 Discussion

The intention of this research was to investigate how substantial cognitive load overshadows the physiological arousal on the PD data (Steinhauer et al., 2004), in a serious game task. More specifically, the investigation was focused on the behavior of the PD in relation to the HR signal. This investigation was based on a challenging decision–making task in a serious game with on–line biofeedback of physiological arousal.

Findings from this experiment suggest that the PD shows the moment when a substantial cognitive load overshadows the physiological arousal effect in relation to HR data, shown in Figure 8.4. This effect increased the PD with physiological arousal until the moment when the dominance occurs, as supported by previous findings (Steinhauer et al., 2004; Stanners et al., 1979). In the context of this study which motivates the mapping of PD to cognitive load and arousal to the HR axis, this would indeed indicate the empirical model of Yerkes–Dodson according to which performance increases with arousal up to a critical arousal point, only to decrease passed that point (Cohen, 2011). Under the assumption of higher performance during periods of increased cognitive load, an optimal level of physiological arousal results in maximum performance on a given task. Taking findings from this experiment together with the previous investigations claiming that a challenging task increases the HR (Boutcher and Boutcher, 2006; Sosnowski et al., 2004) and the PD (Moresi et al., 2008; Recarte and Nunes, 2003), it can be suggested that the PD could be a useful measure of cognitive abilities of participants. However, coupled with another measure (e.g., HR), because of the non–monotonous behavior of the PD across the increasing physiological arousal values. These findings may indicate a potential of serious games to apply physiological metrics to identifying the moment when a substantial cognitive load overshadows the arousal effect, to provide game tasks tailed to the cognitive abilities of an individual interacting with them (Boutcher and Boutcher, 2006). These findings were also supported through the previous research, where it was found that the pupillary response will show an arousal effect only when the cognitive demands of a situation are minimal (Steinhauer et al., 2004; Stanners et al., 1979).

8.6 Limitations

Following each presentation of the price estimation stimulus on the screen, there was a possibility of a pupil diameter change, as well as constriction due to the light reflex effect. Averaging the data over the experiment should have removed these artifacts. Moreover, the fast–paced and highly arousing context of The Auction Game does not allow for the investigation of each individual event in a trial, but an average across a single trial, due to the slow–response heart–rate changes. There might have been a natural physiological delay in the pupillary response followed by the heart response. Therefore, it made sense to average the PD over the period of a heartbeat. The context of the experiment does not allow for the analysis of more short–term measures, due to the nature of the physiological modalities involved, mainly the HR. Therefore, this study focused mainly on the relationship between the PD and HR physiological signals, to establish the possibility of assessing cognitive load from the data. Therefore, the focus wan not on the effect of correct/incorrect decision–making. Furthermore, the serious game context of this experiment might have limited the generalizability of the results towards any decision–making task. Future studies would need to investigate other decision–making task contexts.

8.7 Conclusions

This research contributes to the current body of knowledge by having used the serious game displaying on–line biofeedback of physiological arousal to the participants. Moreover, it successfully identified the moment when a substantial cognitive load overshadowed the arousal effect. These results indicate that the coupling between PD and HR as a measure of physiological arousal may be a useful indicator of cognitive abilities of the participants in the context of a challenging decision–making task. As the task becomes challenging, requiring a substantial cognitive load, dominance of the cognitive load over the physiological arousal effect was observed on the PD. Moreover, HR continued to increase due to the physiological arousal and task difficulty, as supported by previous research (Steinhauer et al., 2004). These findings could provide a step towards understanding the activation underlying the central nervous system mechanisms, and their contribution to cognitive activities. Moreover, it may be beneficial to provide a feedback for the participants to shape awareness of their own cognitive abilities. Usage of physiological signals and observations of the moment when dominance of the cognitive load over the physiological arousal effect occurs could make possible the tailoring of serious games to the cognitive abilities of an individual participant, as previously suggested (Charles et al., 2005).

In summary, this study focused on the research question to find the moment when substantial cognitive load overshadows the arousal effect, in an attempt to investigate cognitive abilities of the participants engaged on a decision–making task. The key contributions of this paper include: identification of the moment when the participants' cognitive abilities are not sufficient to perform on the difficult challenge of the task in the serious game; furthermore, the activation of PNS and SNS have been mapped where their dominance over each–other has been observed, during the course of the task; moreover, the optimal arousal has been identified where maximal performance and the cognitive load have been utilized on the task in the serious game, which is motivated by the reversed 'U' curve relationship between arousal an performance in the Yerkes–Dodson effect (Cohen, 2011).

Future studies are therefore necessary to investigate how a dynamically tailored serious game would influence the participants' cognitive abilities and performance on a decision–making task.

Chapter 9

The Effect of Emotions and Social Behavior on Performance in a Collaborative Serious Game Between Humans and Autonomous Robots

Abstract

The aim of this paper is to investigate performance in a collaborative human-robot interaction on a shared serious game task. Furthermore, the effect of elicited emotions and perceived social behavior categories on players' performance will be investigated. The participants collaboratively played a turn-taking version of the Tower of Hanoi serious game, together with the human and robot collaborators. The elicited emotions were analyzed in regards to the arousal and valence variables, computed from the Geneva Emotion Wheel questionnaire. Moreover, the perceived social behavior categories were obtained from analyzing and grouping replies to the Interactive Experiences and Trust and Respect questionnaires. It was found that the results did not show a statistically significant difference in par-

ticipants' performance between the human or robot collaborators. Moreover, all of the collaborators elicited similar emotions, where the human collaborator was perceived as more credible and socially present than the robot one. It is suggested that using robot collaborators might be as efficient as using human ones, in the context of serious game collaborative tasks.

9.1 Introduction

Collaborative interaction between humans and robots has been a subject of interest for writers and scholars for a long time (Licklider, 1960; Moravec, 2000; Dick, 1968). Motivated by these ideas, researchers working with Human-Robot Interaction (Hri) are investigating robots as peers and colleagues with a variety of social and emotional abilities (e.g., (Norman, 2003) and (Breazeal, 2002)). These endeavors have been motivating Hri research ranging from robotic movement and control, to modeling cognition and social behavior (Fiore et al., 2013). The future might witness an increase in a number of robots, performing tasks in collaboration with humans sharing the same task space (Dautenhahn, 2007; Kanda, 2007; Hinds et al., 2004; Thrun and Bennewitz, 1999). Such scenarios, where humans must work closely with robots in social interaction situations, are becoming increasingly relevant in the world of today (Ros et al., 2014; Cooney et al., 2014; Goetz and Kiesler, 2002; Rehnmark and Bluethmann, 2005; Sidner et al., 2005). This vision gives rise to a need for the investigation of human perception in regards to its robot partners, interacting collaboratively on a shared task. Moreover, it indicates a need for deeper understanding how social cues of a robot collaborator affect performance of a human partner, in an attempt to inform the design of robot collaborators and tasks.

Previous research employing questionnaires shows that embodied robots are sometimes as engaging as humans (Burgoon and Bonito, 2000; Jung and Lee, 2004). A study of perceived social behavior is important for many aspects of robotics, as it investigates perception in Hri and it should be important to consider when designing a human-robot social interaction (Ijsselsteijn et al., 2000). Research shows that engaging physical non-humanoid robot collaborators can convey social cues and be perceived as socially present, furthermore they can elicit emotional responses (Fiore et al., 2013; Scholl and Tremoulet, 2000). People are sensitive to social cues of a robot. In particular, they prefer that a robot moves at a speed slower than that of a walking human (Butler and Agah, 2001; Karwowski and Rahimi, 1991). Moreover, curves of gestured movements influence the perception of collaborating robots (Goodrich and Schultz, 2007). Embodiment characterizes the role of a body in an intelligent behavior (Pfeifer et al., 2012). The authors state that there is a direct link between embodiment and information provided by social cues of robot collaborators. They also state that the human perception of robot collaborators is dependent on the physical interactions, actions, shapes, and the serious game environment itself. The previous research found that social cues (i.e., gestured motion and speed) of collaborating robots elicit emotions in their human partners (Zoghbi et al., 2010). Their study uses variations in gestured motion and speed of the collaborating robot, from a direct path at fixed speed to a gestured path at varying speeds. This investigation is motivated by these findings, as it follows up on the previous study (Zoghbi et al., 2010) also to include perceived social categories, where a robot collaborator acts as a stimulus eliciting emotions and perceived social behavior. Perceived social categories of such interaction (i.e., engagement, reliability, credibility, interaction, and presence) in regards to a robot collaborator, could affect performance on the task presented (Fiore et al., 2013; Burgoon and Bonito, 2002).

Humans use mechanisms from Human-Human Interaction (HHI) to perceive robots as autonomous social agents, more or less as socially present as real human collaborators (Fiore et al., 2013). The previous proposition motivates this research to take into consideration both HHI and HRI for the investigation of perception in social behavior and emotions. This research extends the previous investigation (Fiore et al., 2013) in terms of perceived social presence and it contributes with other social categories and elicited emotions, where these have been brought in relation to performance on a game task.

Previous studies investigating interactions between humans and robots in serious games are sparse. Simulated computer agents playing games such as chess or checkers with or against humans, are familiar concepts (Schaeffer, 1997). However, interaction between humans and robots on a collaborative serious game task within a physical environment is rare. Thus, this investigation aims to extend the research in this domain. To investigate these aspects, this study uses a collaborative serious game task. Although there is a certain popularity of electronic games in current research methods, many traditional games are played in the physical world and require a tangible interaction (Xin and Sharlin, 2007). The physicality of games is important because humans perceive robots as physical entities, capable of interacting with the physical world, while having access to the virtual domain. Such HRI-enhanced traditional serious games can support this physical and virtual duality.

A traditional serious game has been used in the experiment, which can be solved following a sequential set of steps. There exists an optimal solution to the game, such that any non-optimal one could easily be compared for a measurement of human participants' performance. Emerging emotions might be responsible for influencing performance on a game task, as evidence shows that emotions are a key factor that critically influences human decision–making (Loewenstein, 2000; Adam and Gamer, 2011; Picard et al., 2001). For this purpose, the serious game used in this study is turn based and allows one to play at a pace one feels comfortable with, to investigate the effects of robot collaborators' social cues on human performance.

The goal of this paper is to investigate the performance of human participants on a shared serious game task, with a robot partner in their proximate interactive collaboration. In an attempt to investigate how humans perceive robots as social agents and entities in shared-task spaces. In this study, it is of interest to find how a small subset of social cues is perceived and interpreted by humans collaborating with the robot. Thus, this paper aims to investigate the effects of perceived social categories and elicited emotions in proximate interaction, as proposed by (Ijsselsteijn et al., 2000). More specifically, to bring these effects in relation to performance on a serious game task in a collaborative Hri. The contributions of this paper are: 1) to investigate how people perceive robots as social agents in shared-task spaces, capable of interacting with human collaborators, and how it affects performance on a task; 2) understand how social cues of robot collaborators elicit emotions and how those cues are perceived as social categories, affecting the performance of human partners performing collaboratively on s shared game task; 3) understanding both Hhi and Hri in regards to perception of social categories and elicited emotions, bringing results in relation with performance on a game task. These insights could provide a deeper understanding of how emotions and social categories underlie such an interaction from the human collaborator perspective, potentially informing the design of more meaningful collaborative game tasks that would use helpful and intelligent robot collaborators which act according to behavioral patterns that humans can understand and relate to (Xin and Sharlin, 2006).

The remainder of the paper is structured as follows: related work is given in Section 2. Section 3 presents the research questions, while Section 4 details the experimental set-up and the methodological approach to answering them. The results are given in Section 5, while the discussion and the conclusion are given in Section 6 and Section 7 respectively.

9.1.1 Related Work

There is a great body of previous investigations that motivated and supported this research.

9.1.2 Human-Robot Interaction

A Hri requires communication between robots and humans (Goodrich and Schultz, 2007). Authors state that it is “the process of working together to accomplish a goal.” Such communication may take several forms. *Proximate Interaction*, where humans and robots are collocated (i.e., a collaborative robot is in the same room as the humans in this study) (Goodrich and Schultz, 2007). In such interaction it is likely for a robot to be perceived as animate, eliciting social and emotional responses (Fiore et al., 2013). *Social interaction*, where humans and robots interact as peers or companions (Fong et al., 2003). Hri challenges include supporting effective social interactions through social and emotional cues, such as cues associated with the robot's proxemic behavior e.g., gestures and speed. These cues were found to affect significantly the participants' perception of different social categories and emotional states of a collaborating robot (Fiore et al., 2013; Goodrich and Schultz, 2007).

9.1.3 Serious games

Serious games are (digital) games used for purposes other than mere entertainment (Susi and Johannesson, 2007). They need to captivate and engage the player for a specific purpose (Corti, 2006). Social interaction in serious games needs more than one player to be accomplished. Collaboration could be an important method to improve player's motivation, by suggesting serious game activities that involve a physical player to collaborate with on the same task (Hocine and Gouaich, 2013). In traditional collaboration-based digital serious games one is playing together with a computer (a virtual entity). In Hri-enhanced serious games one is playing together with a physical entity eliciting diverse behaviors and stronger emotional responses in players, which might support higher motivation and focus on the task (Hocine and Gouaich, 2013).

Robots were used in serious game settings for treatment of autism (Wainer et al., 2013) and stress (Tapus, 2013), in a game scenario interacting with groups of children (Walters et al., 2005), in games and other natural social interactions with humans conveying emotions and robots providing feedback (Breazeal, 2002; Brooks et al., 2004). Furthermore, robots were used as game companions in an educational setting, where engagement was automatically detected in children playing chess with an "iCat" robot displaying affective and attention cues (Sanghvi et al., 2011). Previous research on interaction between humans and robots in a collaborative serious game within a physical environment, is sparse. The "Sheep and Wolves" is a classic board game, where humans and robots collaborate as independent members of the wolf team. They hunt a single sheep in an attempt to surround it (Xin and Sharlin, 2007). In the "Tic-tac-toe" game the robot and a human can move game pieces on a physical board and collaborate as equals (Tira-Thompson and Halelamien, 2004). Another example is the "Mastermind" game, where the robot makes suggestions to human players as to what colors to pick, where both are engaged in cooperation (Bartneck and Hoek, 2007).

Cue	Description
Engagement	Direct and natural experience
Altruism	Beliefs about a robot's intentions
Reliability	Willingness to depend on the robot
Credibility	How credible is the information
Interaction	How interactive is a robot
Presence	How closely a mediated experience is to an actual, "live" experience

Table 9.1: Concise description of Social interaction cues (Kidd, 2003).

9.1.4 Emotions

Evidence showed that emotions influence decision–making, but they do not always impair it. Emotions were found to have a positive influence on decision–making and to facilitate it (Loewenstein, 2000; Adam and Gamer, 2011; Picard et al., 2001). It was found that the experience of emotions is a mandatory prerequisite for advantageous decision–making (Bechara and Damasio, 2005), therefore having a positive impact on performance (Gross, 1998b; Fenton-O'Creevy et al., 2011; Gross, 2009; Goldstein, 1996). Russell (1980) generally classified emotions by their independent components, arousal and valence. Level of excitement is represented by arousal, while current emotional state is represented by valence, to be positive or negative. Following this notion upon measuring emotions, one is actually measuring a combination of valence and arousal.

9.1.5 Social Interaction

As with other aspects of social interaction, there are many issues that affect perceived engagement, as a number of studies have shown (Burgoon and Bonito, 2000; Koda and Maes, 1996; Reeves and Nass, 1996; Reeves et al., 2003). Robot collaborators were found to be capable of interacting socially with other players (Pereira et al., 2012). Robots would have to possess the skills to interact, as well as to behave in a social manner that would be perceived as acceptable and appropriate in human environments. Table 9.1 gives a concise description of the investigated social categories in this study.

Interaction: The previous studies have investigated robot motion and human perception of robots. For example, the level of comfort for the test subjects was investigated through the Wizard-of-Oz experiments, where an experimenter was hidden and operated or partially operated a robot, thus mimicking autonomous behavior (Koay, 2005; Sisbot and Alami, 2005; Syrdal and Dautenhahn, 2006). Studies have investigated comfort in a Hri (Powers and Kiesler, 2007; Yamato et al., 2003). These studies, specifically investigated social interactions requiring trust and respect, because they were found to be fundamental to many social interactions, including cooperation (Reeves et al., 2003). Studies have shown that people who hold negative attitudes toward robots feel less comfortable in these human-robot situations (Walters, 2005; Syrdal and Dautenhahn, 2006).

Reliability: Goetz et al. (2003) reported that people complied more with a serious, more authoritative robot than with a playful robot on the task that itself was serious. Following this proposition, it seems that the manner in which a robot was presented to humans collaborating with it may have affected the extent to which people were willing to rely on it (Hinds et al., 2004). A dependable robot helped to increase trust in human collaborators, believing that it would be consistent with its operations, and that it would be available when it was needed (Burgoon and Bonito, 2002).

Credibility: How a person perceives robot's motives was found to be important for how credible it was perceived (Burgoon and Bonito, 2002). A desirable robot was the one perceived to have had the person's best interests in mind, and it would often be viewed as beneficial to someone interacting with it. Credibility could have been affected by various causes such as the presence of robots (Burgoon and Bonito, 2002; Reeves et al., 2003).

Engagement: Results show that engagement with robots is dependent on the cues displayed in the game and social context, with prior having a more important role (Castellano et al., 2012). An engaged person was found to have a direct and natural experience, rather than just processing of symbolic data presented (Lombard et al., 2000). Moreover, Sidner et al. (2004) defined engagement as the process of establishing, maintaining and ending an interaction among the partners. Such interaction was therefore found to likely be more compelling.

Presence: Social presence could shortly be described as the "sense of being together with another" (Pereira et al., 2012). It might have an influence on task performance, where impoverished social presence undermines it (Burgoon and Bonito, 2002). If human players perceived artificial opponents as not socially present, their enjoyment would decrease while interacting with them (Heerink et al., 2008). Social presence was found to be important for understanding another agent's intentions (Fiore et al., 2013). In a task where a robot was conveying information to a person (such as giving directions, teaching a new skill and collaborating towards a particular goal), it has been perceived as persuasive with clear intentions (Bengtsson and Burgoon, 1999; Burgoon and Bonito, 2000).

9.2 Research Questions

Following up on the findings and the proposition that the perception of robots as autonomous social agents uses mechanisms from HHI and could affect performance (Fiore et al., 2013), this study applied these mechanisms on the influence of robot collaborators in regards to performance on the game task. Furthermore, this study extends previous research (Fiore et al., 2013) to take into consideration the human collaborator condition. Thus, the following research question is presented:

RQ 1: How do collaborator conditions (human and robot) affect performance in a collaborative serious game?

Research suggests that robot collaborators elicit emotions in human partners (Zoghbi et al., 2010; Goodrich and Schultz, 2007). Research has shown that a high level of focus on a task maximizes performance, while it is correlated with positive emotions and sufficient arousal (Chen, 2007; Csikszentmihalyi, 2008). This is valid unless the challenge is sufficiently beyond or below one's abilities, which generates anxiety or boredom respectively, resulting in worse performance. These studies warrant further investigation on how performance on a collaborative task is affected by the different dimensions of elicited emotions in regards to robot collaborators. To expand on these investigations, the following research question is presented:

RQ 2: How do elicited emotions affect performance in a collaborative serious game between collaborator conditions (human and robot), in regards to valence and arousal?

Research has shown that engaging physical non-humanoid robot collaborators could convey social cues and be perceived as social agents (Fiore et al., 2013; Goodrich and Schultz, 2007). High social presence was found to be correlated with higher performance on the task (Fiore et al., 2013; Burgoon and Bonito, 2002). This study extends on the previous investigations to identify which social categories have an impact on performance in a collaborative task with robot partners. Following these statements, the following research question is proposed:

RQ 3: How do social categories (engagement, reliability, credibility, interaction, and presence) affect performance in a collaborative serious game, between collaborator conditions (human and robot)?

9.3 Methodology

9.3.1 Participants

This study included 70 participants (58 male and 12 female). The age range of the participants was between 19 and 31 with a mean of (23.56 ± 2.338). Subjects were students from The Blekinge Institute of Technology. They were given a movie ticket as a reward for participating. Demographic data were collected regarding the familiarity with the ToH game task, board games in general, and solving mathematical problems.

9.3.2 Compliance with Ethical Standards

Before running a trial involving physical interaction between humans and robots, certain legal and ethical issues must be satisfied. Risk assessment was carried out as suggested in (Crown Copyright, 2003) and mandatory activities were considered (Walters et al., 2005). The experiments were carried out by the Game Systems and Interaction Laboratory (GSIL) at Blekinge Institute of Technology, Karlskrona, Sweden. The Ethical Review Board in Lund, Sweden, has approved all experiments (reference number 2012/737) conducted in this study. The informed consent form was signed by each participant.

9.3.3 Experimental Setup

A crossover study with controlled experiments was conducted in a laboratory setting. Lighting and temperature were controlled in such a way that artificial fixture light was used throughout the experiment while the temperature was held constant at 23°C ± 1°C. The participants were seated in a chair with fixed height and predefined position. The height and position were constant during the experiment. The two experimenters were always present in the laboratory room to monitor the experiments, but they were completely hidden behind the screen. The experimenters were present for safety reasons when using the robot arms and for monitoring that all data were recorded correctly. Surveillance of the robot arms and the participants was done a live feed from a video camera.

In addition, the robot control software included an emergency stopping sequence that would be triggered if defects in the program execution were detected, which was controlled at all times by one of the experimenters. The additional software robot blockage was introduced in the case of any software failures.

Data Collection

The Geneva Emotion Wheel (GEW) was used for self-assessment report of the elicited emotional experience (Scherer, 2005). This tool allowed for fine-grained sampling of a large spectrum of emotional experiences as it includes 20 emotion categories (plus 'other emotion' and 'no emotion' categories), evenly sampling both negative and positive emotions. Arranged in a circle, the tool mapped emotion quality on a two-dimensional valence-arousal space. The emotion intensity was symbolized as the distance from the origin, graphically represented as a set of circles with an increasing circumference (creating an ordinal scale from 0 to 5). The participants were instructed to rate their emotional experience after every run of the game task, by marking the intensity of elicited emotions for which the experience was most salient.

The original Interactive Experiences Questionnaire was developed as a standardized survey for testing presence, specifically for feelings of presence in films (Lombard et al., 2000). The questionnaire was adapted by Kidd and Breazeal (2004) to measure the perceived presence in three characters: a human, a robot, and a cartoon robot. The survey for this study was adapted from Kidd and Breazeal's Interactive experiences questionnaire. Replies to both questionnaires were grouped based on the previous research (Kidd, 2003; Kidd and Breazeal, 2004) (see Appendix A). The replies were given on a seven-point Likert scale, after which they were summed and normalized before the analysis (Likert, 1932).

Regarding the Trust and Respect questionnaire, questions in the first two sections were taken from an earlier questionnaire which was designed to measure five of the six dimensions of social presence (Lombard et al., 2000). The concept of presence as social richness referred to a medium being seen as "sociable, warm, sensitive, personal or intimate when it is used to interact with other people" (Lombard et al., 2000). As in the study from Burgoon and Bonito (2002), responses from the *Solo*, *Human* and *Robot* collaborator conditions were counterbalanced to avoid the ordering effects.

To measure the trust felt by a person towards a robot, the receptivity/trust sub-scale of the Relational Communication Scale was adapted to apply to the robot (Rubin et al., 2010). The adaptation in this case was simply changing "he/she" in the original scale to "the robot" or "the human" (see Appendix B).

Calculations of Cronbach's alpha reported in the appendixes were done on the data that were presented in (Burgoon and Bonito, 2002).

Task Description and Reasoning

The Tower of Hanoi (ToH) serious game was used as a tool for the investigation purposes in the context of human-robot collaborative interaction. The game was easy for the robots to handle since an optimal solution to the game exists, and it was a reasonable challenge for most humans. ToH was originally a single player game. During the collaborative gameplay, human-human or human-robot took turns to complete the game. The rules were explained in (Lucas, 1893). In short, ToH is a mathematical game consisting of three rods on which a number of disks of different sizes can be placed. The goal of the game is to start from a given configuration of the disks on the leftmost peg and to arrive with a minimal number of moves at the same configuration on the rightmost peg (Lucas, 1893). In this study, the serious game started from a given configuration of the disks, which was the same for all participants and referred to the beginning configuration in the game definition above. The individual turns consisted of moving any single disc to a next legal position, interchangeably between the human and robot collaborator until the final configuration of disks was reached on the opposite peg from the start. The human participant always started first. Also, similar to implementing artificial intelligence (AI) for checkerboard games (Xin and Sharlin, 2007), ToH's AI rules were relatively simple. In games where social actions are not required (e.g., chess or checkers) and there exists only one human opponent, AI agents that play optimally are possible (Pereira et al., 2012). At any given moment during the game there was just one possible optimal step to move a disk towards the final configuration. The participants or the collaborators always had an option to select the optimal step as their next move. The participants also had an option to move a disk in any other legal position, which would not necessarily lead towards the final configuration. Such a non-optimal step was only allowed to the participants as the collaborators always took the optimal step to move a disk towards the final configuration.

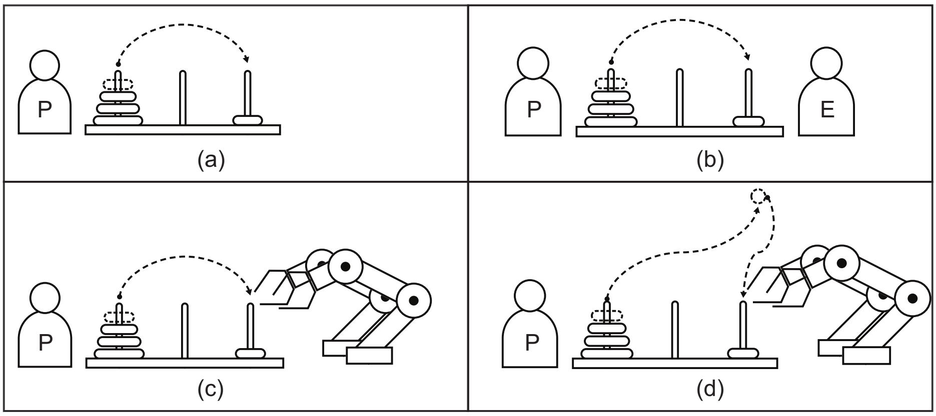

Figure 9.1: Experiment conditions: (a) the *Solo* condition in which a participant is playing the game on his own; (b) playing with the *Human Collaborator* emulating the direct robot collaborator condition, with a direct path at fixed speed; (c) the *Direct Robot Collaborator* condition, always moving in a similar fashion with a direct path at fixed speed; (d) the *Non-Direct Robot Collaborator* condition which had one additional non-direct random point inserted in its path when performing the moves at varying speeds

The effect of human and robot collaborator conditions on performance in a serious game was investigated in this experiment. The main manipulations are presented on Figure 9.1: (a) the *Solo* condition in which the participant is playing the game on his own; (b) playing with the *Human Collaborator* emulating the direct robot collaborator condition, with direct path at fixed speed; (c) the *Direct Robot Collaborator* condition, always moving in the similar fashion with direct path at fixed speed; and (d) the *Non-Direct Robot Collaborator* condition which had one additional non-direct random point inserted in its path when performing the moves at varying speeds. The experiment setup was identical between the trials and participants, where the human participants played the turn-taking ToH game together with a robot or human collaborator, with the same goal of moving the disks from the starting to the final configuration. The obvious exception to this experimental setup was the *Solo* condition, where no collaborators were present and the participants had to play the game on their own. The experimenter had been trained to interact the same way with every participant according to a well rehearsed script, following an optimal algorithm at every move. Participants were instructed on the rules of the ToH game and trained through a practice session with the experimenter until they could finish the simple setup with three disks.

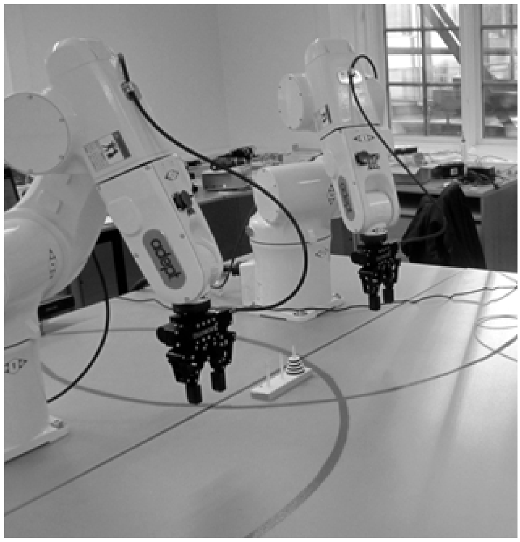

Figure 9.2: Demonstration of the experimental setup where the humans and the robot were collaborating on the ToH serious game, sharing the same physical space

Emotion elicitation and social behavior were achieved by different gestured motions and speeds of the collaborating robot, designed as a platform for the investigation of human players' perception. Concerning robot arms, 51 cm/s for speed of a large and 63 cm/s for a small robot is considered comfortable (Karwowski and Rahimi, 1991). The *direct* gestured motion traced a path between two end points of a current disc movement at a fixed speed of 30 cm per sec (30% of robot speed), see Figure 9.1 (c). The *non-direct* gestured movement traced a path between two end points of a current disc movement, in between which a random point in space above the disks was inserted, randomized on each turn; furthermore, such "random" path was traced at 5 cm/s up to 70 cm/s (5% to 70% of robot speed), also randomized on each turn, see Figure 9.1 (d). Speeds for the *Non-Direct Robot Collaborator* were equally distributed across the exploration space, resulting in the homogeneous distribution of speeds. The robot arm passed through all the three mentioned points in space while making its move before arriving at a final disc movement position. A demonstration of the experimental setup is shown in Figure 9.2. It shows the experimental setup of a human-robot cooperation.

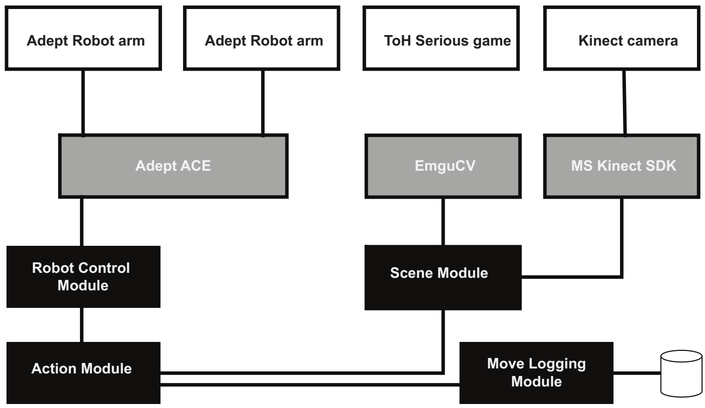

Figure 9.3: The overview of the software system for the ToH serious game. The gray boxes were third-party modules for communication with the hardware, while the black boxes were controller modules for the hardware and the serious game. The experiment data for the analysis were logged to the database (bottom right)

9.3.4 Hardware System

The hardware system (Hagelbäck et al., 2014) contained two Adept Viper S6501[1] 6 DOF robot arms with Robotiq Adaptive 2-finger Grippers[2] as the end effectors. The robot arms were controlled using the Adept SmartController CX control box and two Adept Motionblox-40-60R power adapters. The end effectors were controlled using two Robotiq K-1035 control boxes. The Microsoft Kinect camera was used to monitor the game state by tracking moves made by the humans and robots during the ToH game. A camera was also used for surveillance of the participants and robot arms, in case of an emergency. A single PC running Windows 7 was controlling the system.

9.3.5 Software Platform

The overview of the software system is shown in Figure 9.3. The Action module was the core of the software system. It decided what moves to make and when

[1] `www.adept.com/products/robots/6-axis/viper-s650/general`, accessed 06/10/2017 09:09
[2] `robotiq.com/products/2-finger-adaptive-robot-gripper`, accessed 06/10/2017 09:11

to make them. The Scene module provided information about the moves made and which player was next to make a move. All the moves and the timestamps for when they were made, were saved to the disk using the MoveLogging module. The RobotControl module was responsible for executing moves. It handled the movement of the robot arms to pick up and drop the specified game disk, and when to close/release the grippers. The Scene module used Microsoft Kinect SDK 6[3] to connect to the Kinect camera, and the EmguCV5 (a C# version of OpenCV[4]) library to construct a scene of the current game state. The RobotControl module used the Adept ACE software[5] to control the robot arms. The Data collection software, the robot control software and the Kinect camera software were running on the same computer in order to sync the timestamps between different data files.

9.3.6 Experiment Procedure

The Electrocardiograph, Galvanic Skin Response and Electroencephalograph physiological sensors were present, but they were not used to analyze social behavior categories or elicited emotions reported through the questionnaires in this study.

Upon arrival, the following procedure was employed:

1. After entering the lab room, each participant was seated in a fixed chair at the table and faced the game task at a distance of 50-60 cm.
2. Participants were given written information about the experiment and the description explaining the ToH game. Before starting the experiment session, participants played a practice ToH game with three disks in order to acquaint them with the game task. They were also given written information explaining that the data will be stored confidentially. When the participants agreed to take part in the experiments, they signed an informed consent form.
3. Before the experiment started, participants filled in a demographics questionnaire.
4. Each participant performed an experiment with the four conditions: *Solo*, *Human Collaborator*, *Direct Robot Collaborator* with direct gestured movements and *Non-Direct Robot Collaborator* with non-direct gestured movements. The order of the conditions was counterbalanced between the partic-

[3]`developer.microsoft.com/en-us/windows/kinect`, accessed 06/10/2017 09:14
[4]`opencv.org`, accessed 06/10/2017 09:14
[5]`www.adept.com/products/software/pc/adept-ace/general`, accessed 06/10/2017 09:15

ipants in order to minimize the ordering effects. Each experiment condition was conducted as follows:

(a) A participant played the ToH game.

(b) After a trial was finished, a participant was asked to mark his/her emotional state on the Gew.

(c) The operator instructed a participant about what game task to perform next, by showing information signs on a laptop placed next to the subject. The operator controlled the laptop using a remote desktop.

The four conditions (*Solo*, *Human Collaborator*, *Direct Robot Collaborator* and *Non-Direct Robot Collaborator*) were presented to each participant, as proposed in the study (Burgoon and Bonito, 2002). For each of the four conditions, the participants played the ToH game three times (thus, a total of 12 games were played per participant). Each experimental session took around 90 minutes to complete.

9.3.7 Data Processing

Each reply to the Gew questionnaire was mapped to a point in Russell's valence-arousal space (Russell, 1980), where valence describes if an emotion is negative, positive, or neutral; while arousal describes the physiological activation state of the body ranging from low to high. Replies to the other questionnaires were scored on a seven-point Likert scale, where 1 meant "strongly disagree" and 5 meant "strongly agree".

The reply scores to all questionnaires were transferred into Spss statistics software and analyzed, together with the performance measures consisting of the total number of moves per game and round. Prior to the analysis, the questionnaire data were checked for errors and normalized. Furthermore, outliers were removed using z-scores (standardized values of ± 3.0 or greater).

9.4 Results

Seventy participants from the Blekinge Institute of Technology were randomly allocated to four conditions of solving the ToH serious game. The experiment was counterbalanced such that every participant performed all of the task conditions (*Human Collaborator*, *Direct Robot Collaborator*, *Non-Direct Robot Collaborator* and *Solo*). Reported arousal and valence variables were computed from the

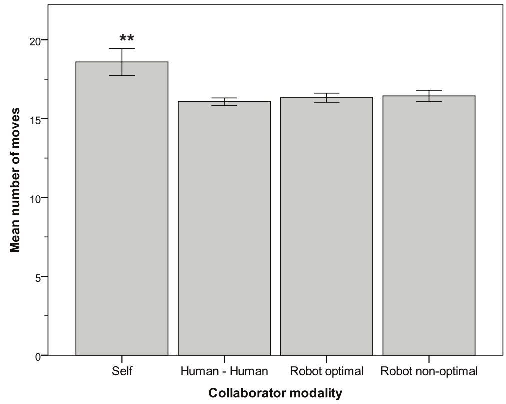

Figure 9.4: The average number of moves per trial in the ToH serious game for each collaborator condition with 95% confidence interval. Stars (***) indicate a significant difference between the *Solo* condition in contrast to any of the collaborators, at the $p < .001$ probability level

Gew questionnaire, while the social category variables (engagement, reliability, credibility, interaction and presence) were obtained from analyzing and grouping replies to the Interactive Experiences (Kidd, 2003; Kidd and Breazeal, 2004) and the Trust and respect questionnaire (Lombard et al., 2000; Rubin et al., 2010), shown in Appendix A and B. Analyses were performed using Spss with the alpha level at 0.05 for all statistical tests. Differences were analyzed using one-way Anova. Correlations were explored using Pearson product-moment correlation index for continuous variables. The data showed no violation of normality, linearity or homoscedasticity (data was normalized and outliers were removed before the analysis). Participants reported previous experience with robotics on a seven-point Likert scale ($\mu = 1.7$ $\sigma = 1.095$ $N = 70$), while 43 of 70 participants have not had any previous experience with the ToH. All of the participants in the experiment were able to finish the trials.

9.4.1 Collaborator conditions and performances

Regarding RQ 1, all of the collaborator conditions were found to have comparable performance scores, as there was no statistically significant difference in the total number of moves per round among them, as determined by a one-way Anova ($F(2,\ 624) = 1.652$, $p > .05$), as presented in Figure 9.4 where higher values reflect worse performance. However, a higher number of moves was expected in the

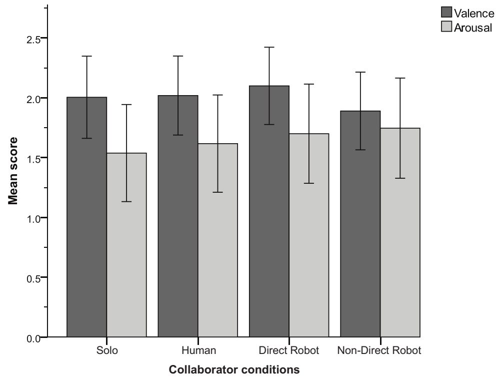

Figure 9.5: The average valence and arousal scores computed from replies to the GEW questionnaire in the ToH serious game for each collaborator condition with 95% confidence interval

Solo condition in contrast to any of the collaborators, as the participants were not expected to know the most optimal solution for the ToH game task. The difference was statistically significant between the groups as determined by one-way ANOVA ($F(3, 824) = 21.215$, $p < .001$). A Tukey post-hoc test revealed that the number of moves to complete the serious game task was statistically significantly lower for the *Human Collaborator* (16.077 ± 1.73, $p < .001$), *Direct Robot Collaborator* (16.382 ± 2.184, $p < .001$), and the *Non-Direct Robot Collaborator* (16.442 ± 2.607, $p < .001$) compared to playing *Solo* (18.597 ± 6.158).

9.4.2 Elicited emotions and performances

The analysis of elicited emotions was conducted to understand their effect on the performance scores (number of moves) on the game task, to answer RQ 2. The serious game presented in this study was reported as arousing, with an average arousal score in the experiment of 1.65 ($SD = 3.017$) of a 10-point nominal scale (GEW) where higher positive values reflect higher arousal. Furthermore, all of the conditions were reported as equally arousing, since there was no statistically significant difference in the reported arousal values between the condition groups, as determined by one-way ANOVA ($F(3, 834) = .194$, $p > .05$), shown in Figure 9.5.

The better performing participants reported higher arousal values after each round, as Pearson product-moment correlation was run to determine the relationship between the participants' number of moves and their reported arousal values in the Gew questionnaire. There was a strong, negative correlation between the total number of moves per round and the reported arousal values in the Gew questionnaire, which was statistically significant (r = -0.094, N = 828, p = .006).

The serious game presented in this study was reported as emotionally positive, with the average valence score in the experiment of 2.0 (SD = 2.424) out of a 10-point nominal scale (Gew), where higher positive values reflect more positive valence. Furthermore, all of the conditions were reported as equally positive. There was no statistically significant difference in the reported valence values between the conditions, as determined by one-way Anova ($F(3, 834) = .266$, $p > .05$), presented in Figure 9.5.

Participants performed the task equally well regardless of reported valence, as no significant correlation was found between reported valence and the number of moves for the *Human Collaborator* (r = -0.109, N = 209, $p > .05$), *Direct Robot Collaborator* (r = -0.079, N = 210, $p > .05$) and *Non-Direct Robot Collaborator* (r = -0.029, N = 206, $p > .05$) conditions.

9.4.3 Social presence and performances

To investigate RQ 3, regarding social presence attributed to the robot collaborators, the analysis to understand the effect of social categories on the game task performance is reported.

The serious game task environment presented in this study was reported with high scores in the social categories during the interaction (the scores are presented as a Likert scale from 1 to 7 with higher values reflecting higher presence of social cues), as shown in Figure 9.6. Furthermore, all of the robot collaborator conditions were found to have equally high scores in the social categories, as reported by the replies to both of the questionnaires. The results showed a significant difference only in the Credibility ($F(2, 207) = 10.341$, $p < .001$) and Presence ($F(2, 207) = 8.751$, $p < .001$) scores, where a statistically significant difference between the conditions was found, as determined by one-way Anova. Tukey post-hoc test revealed that the Credibility score was statistically significantly lower for the *Direct Robot Collaborator* (5.378 ± .889, $p < .001$) and *Non-Direct Robot Collaborator* (5.636 ± .802, $p = .027$) conditions, compared to the *Human Collaborator* one (5.985 ± .672). Tukey post-hoc test revealed a similar finding for the Presence score which was also significantly lower for the *Direct Robot Collaborator* (4.564 ± 1.093, $p < .001$) and the *Non-Direct Robot Collaborator* (4.822 ± .94, $p = .031$) conditions compared to the *Human Collaborator* one (5.23 ± .79). A difference between the human and the robot collaborators

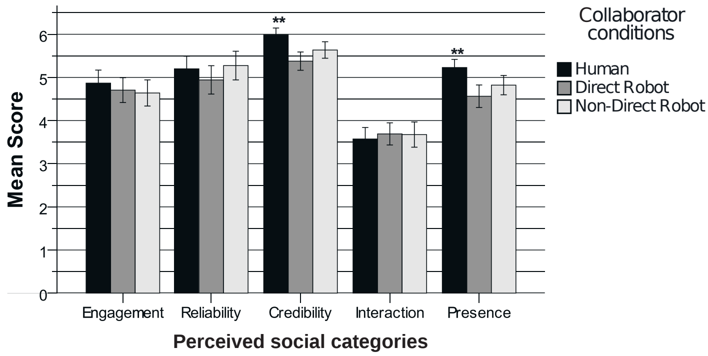

Figure 9.6: The average score for the Perceived social categories in the ToH serious game for each collaborator condition. The scores represent replies to the Interactive Experiences and Trust and respect questionnaire given on a seven-point Likert scale. The replies were grouped and summed for each Social presence cue, normalized and averaged over the whole experiment with 95% confidence interval. The stars (**) indicate a significant difference in the Credibility and Presence scores between the *Human Collaborator* and both of the robot collaborator conditions, at the $p < .01$ probability level

in regards to the Credibility and Presence scores was expected, as humans use mechanisms from HHI to perceive robots as autonomous social agents, more or less as socially present as real human collaborators (Fiore et al., 2013). There were no statistically significant differences between the conditions for the Engagement, Reliability and Interaction ($p > .05$). This finding supports the claim from (Burgoon and Bonito, 2000; Jung and Lee, 2004) that embodied robots are sometimes perceived as engaging as humans.

The participants who performed worse reported a higher social categories score, as Pearson product-moment correlation was run to determine the relationship between the participants' number of moves and their reported questionnaire replies. There was a strong positive correlation between the total number of moves per round and the social categories score for the Engagement, Credibility, Interaction and Presence, which was statistically significant ($r = .300$, $N = 210$, $p < .001$). While social categories score for the Reliability showed no significant correlation to the total number of moves per trial.

9.5 Discussion

The intention of this research was to investigate the performance of participants in a proximate interaction with human and robot collaborators, on a shared serious game task. Furthermore, this study examined the effects of elicited emotions and perceived social categories in the human-robot collaborative performance on a game task. This investigation was based on social cues (i.e., gestured movement and speed) of the collaborating robot. Overall, the collaborators in the ToH serious game elicited arousing, pleasantly valenced emotions and were perceived with high scores in the social categories. The participants performed equally well on the serious game task, regardless of the collaborator condition. Furthermore, both human and robot collaborators were found to elicit emotions which were equally arousing, with equally positive valence. The results further showed that higher arousal scores resulted in higher performance. The participants also reported significantly higher scores in both the Credibility and Presence categories for the human collaborator, compared to both of the robot collaborators.

Considering RQ 1, findings from this experiment regarding the performance on a serious game task suggested that collaborating with a robot collaborator might have been as effective as collaborating with a human one. Results from the experiment did not show a statistically significant difference in the participants' performance between any of the collaborator conditions. This may have been because participants were highly focused on the game task, since collaborating with a physical entity eliciting diverse behaviors and strong emotional responses might have promoted a higher focus on the task (Hocine and Gouaich, 2013; Chen, 2007; Csikszentmihalyi, 2008). As expected, worse performance was found in the *Solo* condition, possibly because help was not available from the collaborators, leaving the participants to their own skills and with enough room for non-optimal moves. All of the collaborators were playing optimally on each move.

The relevance of previous claims was further supported through RQ 2, exploring the performance in regards to elicited emotions on a collaborative serious game task. The results indicated that both the human and robot collaborators elicited emotions which were equally arousing, with equally positive valence in the context of the ToH serious game. This implied that the investigated social cues and the human collaborator condition have not had a significant effect on the elicited emotions. These findings were also supported through the previous research, where participants collaborated with physical entities eliciting diverse behaviors and strong emotional responses (positive emotions and sufficient arousal), which might have supported a higher focus on the task (Hocine and Gouaich, 2013; Chen, 2007; Csikszentmihalyi, 2008). Furthermore, the results showed that higher arousal scores resulted in higher performance while valence had no significant effect, especially since these findings hold true even for the *Solo* condition in this investigation. These findings were supported through the previous research investigations, where high levels of focus on a task were correlated with positive emotions and sufficient arousal, which might have increased the performance on the task (Chen, 2007; Csikszentmihalyi, 2008). These results provided further evidence that interacting with a robot collaborator might have been as effective as collaborating with a human one, in regards to eliciting comparable levels of arousal and valence.

Regarding RQ 3, the investigation of the performance in regards to the diverse social categories perceived suggested that, the human collaborator was perceived as significantly higher in both the Credibility and Presence categories, compared to both of the robot collaborators. This difference had been supported by the previous research, where humans use mechanisms from Hhi to perceive robots as autonomous social agents, more or less as socially present as real human collaborators (Fiore et al., 2013), where human collaborators were found to be perceived as more present and credible than robot collaborators. Taking these findings into consideration, this study showed a similar performance in the participants collaborating with both the human and robot collaborators. These findings were comparable with findings in the previous studies, where similarly perceived robot collaborators resulted in similar performance regarding social presence (Fiore et al., 2013). Moreover, those social categories could have supported the design of embodiment in robots, as embodiment is reciprocal and dynamical coupling between brain (control), body, and environment (Pfeifer et al., 2014). This implied that further investigations in human-human collaboration is needed to map out the social cues responsible for this difference, in an attempt to design more meaningful robot collaborators and serious games. Furthermore, the worse performing participants reported higher scores on all of the perceived social categories, except in the Reliability, where no significant correlation was found. A possible explanation could be that the participants who performed worse had a more meaningful and longer interaction with the collaborators, which potentially could have resulted in higher reported scores.

9.5.1 Limitations

The limitations of this study were several. The collaborative serious game context of this experiment and the sole serious game involved, might have limited the generalizability to any collaborative task between humans and robots. Therefore, future studies would need to include other task and serious game contexts to present a convincing general conclusion. Furthermore, the human collaborator was in a position to express other social cues that were not investigated in this study through the robot collaborators. Future studies should investigate further into the social cues of human-human collaboration. The Trust and Respect Questionnaire might not have been exhaustive enough to describe the social presence. Therefore, when findings regarding the social presence were presented, they were regarded as a sense of working together with another to accomplish a goal (Pereira et al., 2012; Goodrich and Schultz, 2007), as previously used in the literature (Lombard et al., 2000). The sample population of the participants was entirely made of college undergraduates and graduates. As social robots will be integrated in various contexts, future research should deepen the understanding of different populations interacting with robot collaborators. Furthermore, there might have been a relatively weak manipulation of the two intended social cues, as these had no significant influence on performance on the task.

There might have been fatigue effects present in this experiment, but it was assumed that they were minimal. Even more so, since the results showed an increased performance even during the later trials for each participant. The experiment lasted for around 90 minutes, with the time on the task around 30 minutes, which was comparable to other similar experiments reported.

9.6 Conclusion

This research contributes to the current body of knowledge by having used a realistic setting in which participants actually interacted with an autonomous robot collaborator to solve a serious game task. This stands in contrast to the previous Wizard-of-Oz experiments, where an experimenter was hidden and operated or partially operated a robot, thus mimicking an autonomous behavior (Koay, 2005; Sisbot and Alami, 2005; Syrdal and Dautenhahn, 2006). Furthermore, questionnaire data were collected continuously throughout a session, directly after the game task had ended (the moment of the actual experience). Such reports might have been more accurate than the reports collected after the whole session or experiment, as a recollection of an experience. Additionally, the reporting was automated, thus not having had an experimental environment contaminated with anybody other than the participants, human and robot collaborators.

Overall, the collaborators in this serious game elicited arousing, pleasantly valenced emotions and were perceived with high scores in the social categories. The results of this experiment indicated that participants' performance on the serious game task is comparable between the human and robot collaborator conditions. Even more so, as all of the collaborator conditions elicited similar levels of emotional arousal and valence, which might have been beneficial for having a higher performance on the task. These findings motivate the introduction of autonomous robots as collaborators on tasks. Moreover, the (non-humanoid) robots and human collaborator in this study elicited similar positive emotional valence in the participants. This is important, as robots get introduced in different aspects of human lives, possibly as team members (Fiore et al., 2013). On the other hand, the human collaborator was perceived as more credible and socially present, but it had no effect on the actual performance on the task as there was no significant difference in participant's performance between the collaborator conditions. Current results support the notion that understanding emotional and social cues underlying such collaboration from the human perspective, it would be possible to design more helpful and intelligent robots which act according to behavioral patterns that humans can understand and relate to (Xin and Sharlin, 2006). Especially if we consider humans who will have a longer and more meaningful interaction with robot collaborators, as this study found that those individuals reported higher scores on all of the perceived social categories, except on the Reliability. Such investigations would seek to improve the quality of HRI by designing autonomous robot collaborators and serious games, that will rely on perceived social behavior and elicited emotions in an attempt to be more natural, intuitive and familiar for its users (Breazeal, 2002). If one also considers that robots possess physical-virtual duality and have access to the game task information, one can clearly see why a choice of robot collaborator on a task would be a sound choice.

Nevertheless, in the context of a serious game collaborative task, findings from this study supported the notion that using robot collaborators might be as efficient as using human ones. The future might witness the design of robot collaborators as socially closely perceived as their human counterparts, in regards to social presence and engagement. Especially since no significant difference was found in the participants' performance interacting with the human collaborator. Future studies should investigate further into social cues in HHI and HRI for other serious game contexts, to inform a more general design of robot collaborators, which would be as credible and as socially present as human ones. Furthermore, future research should also investigate participants' emotions and their detection using the Psychophysiology as an important part of HRI, where the embodiment is a powerful concept in the development of adaptive autonomous systems (Pfeifer et al., 2014). As proposed in (Pereira et al., 2012), these future studies could bring about an increase in perceived social categories of robot collaborators and facilitate a more meaningful social interaction with their partners.

Acknowledgments. This work was supported by the EU Project EC FP7-ICT-231143 ECHORD.

The authors are grateful for the support and suggestions of our many colleagues. Special thanks to Craig Lindley and Stefan Johansson for their guidance and suggestions. Also to the other members of the CogNeuro group at the Bleking Institute of Technology, including Olle Hilborn and Johan Svensson, without which this project would not have been possible. Finally, we would like to thank the reviewers for their valuable insights and comments.

9.7 Appendix A - Interactive Experience Questionnaire

Likert scale was used for every question depending on the collaborator modality
Robot: Not at all 1 2 3 4 5 6 7 Very much
Human: Not at all 1 2 3 4 5 6 7 Very much

1. How engaging was the interaction?
2. How relaxing or exciting was the experience?
3. To what extent did you experience a sensation of reality?
4. How much attention did you pay to the display devices or equipment rather than to the interaction?
5. How often did you feel that the character was really alive and interacting with you?
6. How completely were your senses engaged?
7. How natural was the interaction with the character?
8. The experience caused real feelings and emotions for me.
9. I was so involved in the interaction that l lost track of time.
10. How often did you want to or did you move your body or part of your body either closer to or farther away from the characters you saw/heard?
11. To what extent did you feel you could interact with the character?
12. How often did you have the sensation that the character could also see/hear you?
13. How much control over the interaction with the character did you feel that you had?
14. How often did you make a sound out loud (e.g., laugh, speak) in response to someone you saw or heard in the interaction?
15. How often did you smile in response to the character?

16. How often did you want to or did you speak to the character?
17. How often did it feel as if the character was talking directly to you?
18. He/she is a lot like me.
19. I would like him/her to be a friend of mine.
20. l would like to talk with him/her.
21. If he/she were feeling bad, l'd try to cheer him/her up.
22. I looked at him/her often.
23. He/she seemed to look at me often.
24. He/she makes me feel comfortable, as if I am with a friend.
25. I like hearing his/her voice.
26. If there were a story about him/her in a newspaper or magazine, I would read it.
27. I like him/her.
28. I'd like to see/hear him/her again.

9.8 Appendix B - Trust and Respect Questionnaire

(Lombard et al., 2000; Rubin et al., 2010)

Engagement: (Calculated Cronbach's alpha: 0.85)

Robot: Not at all 1 2 3 4 5 6 7 Very much
Human: Not at all 1 2 3 4 5 6 7 Very much

- How engaging was the interaction?
- How relaxing or exciting was the experience?
- How completely were your senses engaged?
- The experience caused real feelings and emotions for me.
- I was so involved in the interaction that l lost track of time.

Reliability: (Calculated Cronbach's alpha: 0.85) All items were ranked on a seven-point Likert scale with a range of "strongly disagree" (1) to "strongly agree" (7).

- I could depend on this robot to work correctly every time.
- The robot seems reliable.
- I could trust this robot to work whenever I might need it.
- If I did the same task with the robot again, it would be equally as helpful.

Credibility: (Reported Cronbach's alpha: 0.90) Rate the robot (Likert scale 1 -7 was used):

- Kind to Cruel
- Safe to Dangerous
- Friendly to Unfriendly
- Just to Unjust
- Honest to Dishonest
- Trained to Untrained
- Experienced to Inexperienced
- Qualified to Unqualified
- Skilled to Unskilled
- Informed to Uninformed
- Aggressive to Meek
- Emphatic to Hesitant
- Bold to Timid
- Active to Passive
- Energetic to Tired

Presence: Please give your impression of the robot: Describes poorly 1 2 3 4 5 6 7 Describes well

- Annoying

- Balanced
- Compelling
- Convincing
- Credible
- Enjoyable
- Entertaining
- Fair
- Favorable
- Good
- Helpful
- Honest
- Homogeneous
- Informative
- Likable
- Negative
- Persuasive
- Reliable
- Satisfying
- Trustworthy
- Useful
- Varied
- Well-composed

Chapter 10

Physiological Affect and Performance in a Collaborative Serious Game between Humans and an Autonomous Robot

Abstract

In this paper, elicited physiological affect in human participants collaborating with the robot partners was investigated to determine its influence on the decision performance in serious games. The participants collaboratively played a turn–taking version of the Tower of Hanoi, where physiological arousal and valence underlying such human–robot proximate collaboration were investigated. Galvanic skin response and electrocardiography were used to infer the physiological affect. Higher physiological affect was found in players collaborating with the robot collaborators, compared to the human one. Furthermore, a comparable decision performance in the serious game was found between all the collaborator conditions. It is suggested that serious games which are carefully designed to take into consideration the elicited physiological affect might witness a better decision performance using robot partners instead of human ones.

10.1 Introduction

An interest of Human-Robot Interaction (Hri) lies in the investigation of robots and their emotional abilities through an interaction with peers or colleagues (Norman, 2003; Breazeal, 2002). In Hri, humans and robots interact together in a shared physical space aimed at accomplishing a goal (Goodrich and Schultz, 2007). Such *proximate interaction* includes factors (e.g., gaze, expressions, gestures, speed, distance) which are perceived to elicit affect in humans, and used to attribute emotional states to robots (Fiore et al., 2013). Numerous studies indicate the importance of *social interaction* between humans and robots (Dautenhahn, 2007). The social collaborative–interaction situations where humans and robots work together in shared task spaces, are gaining a lot of interest (Ros et al., 2014; Hinds et al., 2004). Findings from these studies give rise to a need for the investigation of elicited human physiological affect in regard to their robot partners, interacting collaboratively on a shared task. Moreover, it motivates a need for deeper understanding how elicited physiological affect influences the decision performance of human partners, in an attempt to inform the design of robot collaborators and to optimize the decision performance on collaborative tasks.

The Theory of Mind reasons that humans perceive and distinguish various social cues to explain events in terms of intents and goals of agents (i.e., their actions (Fiore et al., 2013; Leslie, 1994)) which might have an effect on the elicited emotions (De Sonneville et al., 2002). Following these propositions from Human–Human Interaction (Hhi) and Hri, this study set out to understand these mechanisms of ability to create an internal model of another agent based on the observable social cues leading to a better interaction and understanding of the social exchange that occurs between human and robot agents (Fiore et al., 2013; Gallagher, 2007). Apart from the human agents, in the context of this study a robot agent is facilitated as an autonomous social entity that can feature diverse behavior (i.e., completely passive or autonomously acting based on a complex algorithm). The embodiment describes the role of a body in an intelligent behavior (Pfeifer et al., 2012), where a direct link exists between the embodiment and information provided by the social cues of collaborating robots. Humans perceive collaborating robots depending on the physical interaction, actions, shape and the environment itself. Embodied robots have been found to be as engaging as humans (Burgoon and Bonito, 2000; Jung and Lee, 2004). Furthermore, engaging physical non–humanoid robot collaborators have been found to elicit emotional responses (Fiore et al., 2013).

As evidence shows that emotions critically influence human decision–making and performance (Chanel, 2009; Shiv et al., 2005; Gross, 1998b), this study sets out to find how a small subset of social cues elicits physiological affect in humans collaborating with robots, in an attempt to investigate how these influence performance on a serious game task. Humans use the mechanisms from Hhi to perceive robots as autonomous social agents (Fiore et al., 2013). These propositions motivate this investigation to take into consideration both Hhi and Hri, in the investigation of the elicited physiological affect. This research builds upon our previous investigation where autonomic non–humanoid robots were percieved as social and emotional agents using a small subset of social cues in a serious game (Jerčić et al., 2018b). Therefore, this research aims to contribute with a detection method for elicited physiological affect. Moreover, these concepts have been brought in relation to the performance on a serious game task.

Many traditional games have been played in the physical world and require a tangible interaction, in contrast to a certain popularity of electronic games in the current research methods (Xin and Sharlin, 2007). Humans perceive robots as physical entities capable of interacting within the physical world while having access to the virtual domain, thus the physicality of serious games is important. This physical and virtual duality is supported through such Hri–enhanced traditional serious games. This study uses a traditional game which provides an easy measurement of performance through a sequential set of steps following an algorithm. Taking this finding into consideration, the serious game in this study is turn–based and allows for a pace participants feel most comfortable with, to investigate the effects of physiological affect on human performance, elicited through the collaborators' social cues.

The purpose of this study is to investigate the performance of human participants, collaborating with robot partners in their proximate interaction on a shared serious game task. This paper attempts to investigate the physiological affect underlying such human–robot proximate collaboration. Moreover, it aims to understand its role in the performance by mapping the participants' physiological responses towards the collaborating robot on the arousal/valence axes (Pereira et al., 2012). Following this goal, this paper sets out to investigate the effects of the elicited physiological affect on proximate interaction, as proposed by (Ijsselsteijn et al., 2000). More specifically, to bring these effects in relation to the performance on a serious game task in collaborative Hri. The contributions of this paper are: 1) investigate the performance of human participants performing on a collaborative serious game task together with the human and robot collaborators in a shared space; 2) understanding how the elicited physiological affect influences the performance of human participants performing collaboratively on a shared serious game task together with the human and robot partners; 3) understanding of both Hhi and Hri in regard to how the human and robot collaborators elicit physiological affect, bringing these results in relation to the performance on a serious game task. These insights could provide a deeper understanding of how elicited physiological affect underlies such an interaction from the human collaborator perspective, informing the design of more meaningful collaborative serious games that would use the objective measures of physiological affect together with intelligent robot collaborators to potentially increase performance.

The remainder of the paper is structured as follows: the related work is given in Sect. 2. Sect. 3 presents the hypotheses, while Sect. 4 details the experimental set–up and the methodological approach to addressing them. The results from this experiment are given in Sect. 5, while the discussion and the conclusion are given in Sect. 6 and Sect. 7 respectively.

10.2 Related Work

Evidence showed that emotions influence human decision–making and performance (Loewenstein, 2000; Adam and Gamer, 2011; Gross, 2009; Picard et al., 2001). Various emotional models have been reported in the literature, ranging from the emotions viewed as discrete constructs (Ekman, 1999) to their determination by a combination of factors (Schachter and Singer, 1962). Russell (Russell, 1980) classified emotions through a combination of their independent components, arousal and valence. In his model, the level of excitement has been represented by arousal, while valence defined whether the current emotional state is positive or negative. Using this classification one is actually measuring a combination of valence and arousal while measuring emotions. The somatic marker theory claims that decisions are aided by emotions in the form of bodily states, which may contain relevant information (Naqvi et al., 2006; Bechara and Damasio, 2005).

The evidence further suggests that people are sensitive to the social cues of collaborating robots (Goodrich and Schultz, 2007) (i.e., gestured motion and speed (Butler and Agah, 2001; Karwowski and Rahimi, 1991)), where they have been found to elicit emotions in their human partners (Zoghbi et al., 2010). The collaborators in this study act as a stimulus eliciting emotional responses, related together with the performance on a serious game task. This notion is motivated through the previously mentioned investigation (Zoghbi et al., 2010) where variations in gestured motion and speed of the collaborating robots have been used, from a direct path at fixed speed to a variable speed in gestured motions. Therefore, this investigation also includes the human and robot collaborators.

10.2.1 Affect Detection

There is a rich body of physiological literature related to the affect detection and its use in the Hri domain (Bethel et al., 2007). However, a few studies have been using robots as the primary elicitors of affective physiological responses. While the subjective measures (e.g., self–evaluation) always incorporate a potential self–deception, physiological signals are seen as an objective measure of emotional states since they are hard to manipulate intentionally (Jerčić et al., 2012). There is solid evidence that physiological activities associated with affective states can be differentiated and systematically organized, which would allow for the analysis of their effect on the performance in collaborative serious games (Bradley and Lang, 2000).

Electrocardiography (Ecg) and *Galvanic Skin Response* (Gsr) were employed to measure emotional states of humans in response to robots with more than 80% success rate (Rani et al., 2007; Kulic and Croft, 2006; Picard et al., 2001). Studies have found a strong correlation between Ecg and physiological valence (Xu et al., 2010; Andreassi, 2007). There have been multiple findings of a linear correlation between Gsr and physiological arousal (Braithwaite and Watson, 2013; Dawson et al., 2007). Even though Gsr has been found to indicate physiological valence, the correlations were not always as strong, and a greater confidence in interpretations has been obtained by combining Gsr with other measures, such as Ecg (Andreassi, 2007).

10.2.2 Serious Games

There are studies investigating the interaction between humans and robots in serious games, but they are sparse. Even though research on interaction in a physical environment is rare, simulated computer agents playing games (e.g., chess or checkers) together with humans are more common (Schaeffer, 1997). Physical collaborators might support higher motivation and better performance in contrast to the traditional collaboration–based digital serious games (Kiesler and Hinds, 2004), especially for the robot collaborators in serious games (Hocine and Gouaich, 2013). A collaborative serious game task has been used in this study to extend the research in this domain.

Serious games are defined as games which are not used for mere entertainment purposes (Susi and Johannesson, 2007). They need to captivate and engage players for a specific purpose intended (Corti, 2006). In contrast to the traditional collaboration–based digital serious games where one is playing together with a computer (a virtual entity), Hri–enhanced serious games present a physical entity eliciting diverse behaviors and stronger emotional responses in participants (Hocine and Gouaich, 2013). The authors further state that this might support higher motivation, performance and focus on the task (Hocine and Gouaich, 2013; Kiesler and Hinds, 2004). A collaborative serious game task has been used in this study to extend the research in this domain.

The first step in discussing the interaction between humans and robots in collaborative serious games within a physical environment must be a consideration of decision–making on the same task. There have been a number of previous studies which have used robot collaborators aiding human decisions in serious game settings. Robots were successfully introduced and used in a social interaction context, inside serious games: the “Tic–tac–toe” serious game where the robot and humans move the game pieces on a physical board (Tira-Thompson and Halelamien, 2004); for the treatment of autism (Wainer et al., 2013) and stress (Tapus, 2013); and in games and other natural social interactions with humans conveying emotions and robots providing feedback (Brooks et al., 2004). The dynamic that we see between similar performance and higher positive valence for the robot collaborators, in comparison to human ones, will provide a new way to the limited understanding of the human collaboration with robots in serious games, through the detection of underlying emotional states (Agrawal et al., 2008). It is only when we perceive physiological signals in a broader sense that we can grasp their role in understanding physiological measurements of elicited affect using social cues in the Hri–enhanced serious games. Peck et al. (2014) used a robotic basketball game to learn the preferences of children with Autism Spectrum Disorder, where the anxiety levels as affective states were implicitly measured using physiology. The authors reported an increased performance and higher positive valence between the robot collaborator conditions. Kim and Kwon (2010) designed a game of “Twenty Questions” as an interactive task with the robot system successfully detecting emotions. Tapus (2013) created a serious game called the “Stress Game” to elicit stress and consequent frustration in the player, using the robot which continuously monitors player’s performance level to frustrate the player. The authors reported an increased performance for the lower arousal condition.

10.3 Hypothesis

The previous investigations have shown that engaging physical non–humanoid robot collaborators can elicit emotional responses (Fiore et al., 2013; Scholl and Tremoulet, 2000), which in turn might influence human decision–making and performance on a game task (Chanel, 2009; Adam and Gamer, 2011; Shiv et al., 2005). Furthermore, as human collaborators are sensitive to social cues (Goodrich and Schultz, 2007) (i.e., gestured motion and speed (Butler and Agah, 2001; Karwowski and Rahimi, 1991)), they use the mechanisms from Hhi to perceive the observable social cues which might have an effect on elicited emotions (Fiore et al., 2013; De Sonneville et al., 2002). Following up on these findings, this study investigated the influence of human and robot collaborators on the performance in a serious game. Thus, this study extends the previous research (Fiore et al., 2013) to take into consideration the human collaborator condition, taking into account the elicited physiological affect in response to the human and robot collaborators. Therefore, the following hypothesis is presented:

Hypothesis 1: The collaborator condition (i.e., human, robot) will affect the performance on the game task (H1).

Previous research suggests that robot collaborators elicit physiological emotions in human partners (Zoghbi et al., 2010; Goodrich and Schultz, 2007). Moreover, a higher motivation (Hocine and Gouaich, 2013), positive emotions and sufficient arousal (Chen, 2007; Csikszentmihalyi, 2008) are correlated with higher performance in serious games with physical collaborators. Csikszentmihalyi and Bosse (Csikszentmihalyi, 2008) argue that previous proposition is valid unless a challenge is sufficiently beyond or below one's abilities, which might generate anxiety or boredom respectively, resulting in a lower performance. These studies warrant further investigation to how the performance on a collaborative task is affected by different dimensions of elicited physiological affect in regard to robot collaborators.

It has been found that an engaging environment elicits high physiological arousal, while those higher levels of arousal were correlated with a higher performance (Burgoon and Bonito, 2002; Pearson, 2007). To expand on these investigations, this study manipulated the collaborator conditions by varying the social cues eliciting physiological arousal, through speed and gestured motions. Following on the previous findings (Burgoon and Bonito, 2002; Pearson, 2007) that the conditions of proxemic collaborators would influence elicited physiological arousal in participants and therefore performance on a game task, the following hypothesis is presented:

Hypothesis 2: Elicited physiological arousal will be affected by the collaborator condition (H2a), which in turn will affect the performance on the game task (H2b).

While physical collaborators have been found to elicit a higher motivation on a task, which is correlated with positive emotions (Chen, 2007; Csikszentmihalyi, 2008), previous studies have not found a strong correlation between physiological valence and performance (Reeves and Nass, 1996; Burgoon and Bonito, 2002).

To expand on these findings, the following hypothesis is postulated for the investigation of the interactive effects of elicited physiological valence and its influence on the performance in a serious game:

Hypothesis 3: Elicited physiological valence will be affected by the collaborator condition (H3a), which in turn will have no significant effect on the performance on the game task (H3b).

10.4 Methodology

10.4.1 Participants

This study included 70 participants, 58 were males and 12 females. The age of participants ranged between 19 and 31, with a mean of (23.60 ± 2.34). Demographic data (i.e., familiarity with the ToH game task, board games in general, and solving mathematical problems) were collected and they were given a movie ticket as a reward for participating. Participants were students of Blekinge Institute of Technology. The experiments were carried out by the Game Systems and Interaction Laboratory (GSIL) at Blekinge Institute of Technology, Karlskrona, Sweden. The Ethical Review Board in Lund, Sweden, has approved all experiments (reference number 2012/737) conducted in this study.

As explained in our previous investigation (Jerčić et al., 2018b), a crossover study with controlled experiments has been conducted in a laboratory setting. The lighting and temperature conditions were controlled in such a way that artificial fixture light was used throughout the experiment while the temperature was held constant at 23 °C ± 1 °C. The participants were seated in a chair with a fixed height and a predefined position. The height and position were constant during the experiment. The two experimenters were always present in the laboratory room to monitor the experiments, but they were completely hidden behind the screen. The experimenters were present for safety reasons when using the robot arms and for monitoring that all data were recorded correctly. Surveillance of the robot arms and the participants was done using a live feed from a video camera. In addition, the robot control software included an emergency stopping sequence that would trigger if defects in the program execution were detected, which was controlled at all times by one of the experimenters. An additional fail–safe software was introduced in the case of possible software failures.

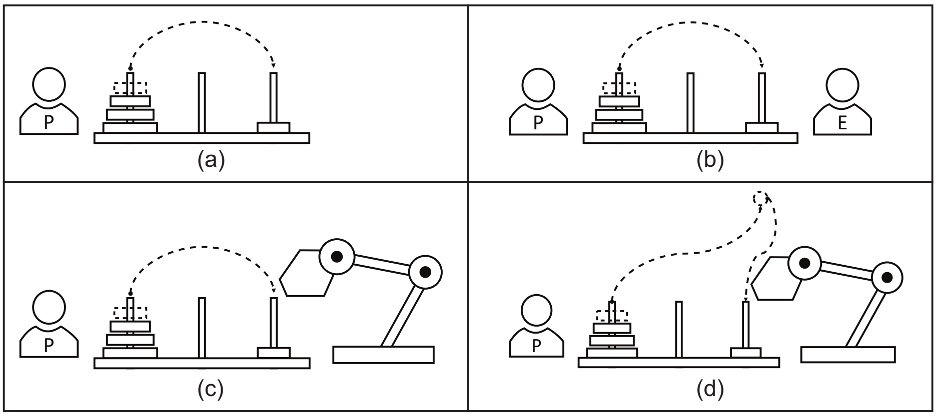

Figure 10.1: Experiment conditions (Jerčić et al., 2018b): (a) the *Solo* condition in which the participant is playing the game on his own; (b) playing with the *Human Collaborator* emulating the direct robot collaborator condition, with direct path at fixed speed; (c) the *Direct Robot Collaborator* condition, always moving in a similar fashion with a direct path at fixed speed; (d) *Non–Direct Robot Collaborator* condition which had one additional non–direct random point inserted in its path when performing the moves at varying speeds.

10.4.2 Experimental Setup

Main Manipulations

The effect of physiological affect on the performance in the serious game, in regard to the human and robot collaborator conditions, was investigated in this experiment. The main manipulations (see Figure 10.1) were: (a) the *Solo* condition where no collaborators were present and in which a participant was playing the game on his own; (b) the *Human Collaborator* condition emulating the direct robot collaborator condition, with a direct path at fixed speed; (c) the *Direct Robot Collaborator* control condition, where the participants played together with the robot which was always moving in a similar fashion with a direct path at fixed speed; (d) the *Non–Direct Robot Collaborator* condition which had one additional Non–Direct random point inserted in its path moving at varying speeds. The experiment setup was identical between the trials and participants, where human participants played the turn–taking ToH game together with a robot or human collaborator. The goal of the game task was to move the disks from the starting to the final configuration. The collaborators were playing optimally on each move, following the algorithm. The experimenter has been trained to interact the same way with every participant according to a well rehearsed procedure.

Study Stimuli

The Tower of Hanoi (ToH) game was used as the serious game for the study to investigate the elicited human physiological affect in collaborative Hhi and Hri, and bring it in relation to the performance on a collaborative serious game task. Most of the participants were naive to the ToH serious game. The ToH was employed as a serious game in this scientific exploration of human–robot collaborative interaction, following the definition of serious games as games with a clear purpose other than pure entertainment (Susi and Johannesson, 2007). The game was easy for the robot to handle since an optimal solution to the game exists, and it was a reasonable challenge for most of the participants. ToH was originally a single player game. In a collaborative gameplay, human-human or human–robot took turns to complete the game. The rules were explained in (Lucas, 1893). In short, the ToH is a mathematical game consisting of three rods and a number of disks of different sizes that can slide onto any rod. The goal of the game is to start from a given configuration of the disks on the leftmost peg and to arrive in a minimal number of moves at the same configuration on the rightmost peg (Lucas, 1893). In this study, the serious game started from a given configuration of the four disks, which was the same for all participants and referred to the beginning configuration in the game definition above. The individual trials consisted of moving any single disc to a next legal position, interchangeably between a participant and a collaborator until the final configuration of disks was reached on the opposite peg from the start. The participants always started first. Furthermore, similar to checkerboard games implementing rules for the ToH was relatively simple (Xin and Sharlin, 2007). In games where social actions are not required (e.g., chess or checkers) and there exists only one human participant, autonomous agents that play optimally are possible (Pereira et al., 2012). At every move in the ToH game, just one possible optimal step existed to move a disk towards the final configuration. The participants always had an option to take this optimal step as their next move, while it was mandatory for the collaborators. Therefore, only the participants had an option to take the non–optimal step to move a disk in any other legal position, which would not necessarily lead towards the final configuration.

The elicitation of physiological affect was achieved by the gestured motions and the speed of the collaborating robot. In particular, humans prefer that a robot moves at a speed slower than that of a walking human (Butler and Agah, 2001; Karwowski and Rahimi, 1991). The gestured motions were composed from a direct path at fixed speed of 30 cm/sec (30% of robot speed) between the two end points of a current disc movement, for the *Direct Robot Collaborator*, to a random path and speed between 5 cm/sec up to 70 cm/sec (5% to 70% of robot speed). A random path and speed were generated online for the *Non–Direct*

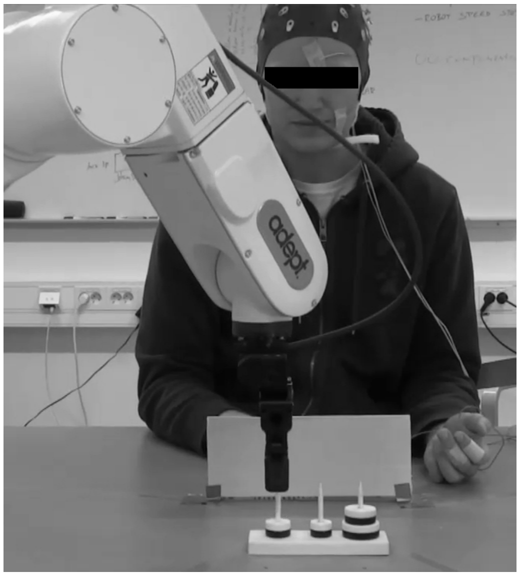

Figure 10.2: Demonstration of the experimental setup where a human and the robot are collaborating on the ToH serious game, sharing the same physical space.

Robot Collaborator. A Non–Direct path is generated using the two end points of a current disc movement in between which a random point in space above the disks was inserted, randomized on each game move robot arm makes, which totals in three virtual positions the robot arm has to follow while making its move. The robot was passing through all the specified positions before having arrived at a final disc movement position.

10.4.3 Experiment Procedure

A demonstration of the experimental setup is shown in Figure 10.2. It shows the experimental setup for a human–robot cooperation with all the physiological sensors attached.

Upon arrival of a participant, the following procedure was employed:

1. After entering the lab room, each participant was seated in a fixed chair at the table and faced the game task at a 60 cm distance.

2. The participants were given written information about the experiment and the description explaining the ToH game. Before starting the experiment session, the participants played a practice ToH game with three disks in order to acquaint them with the game task. They were also given written information explaining that data were stored confidentially. When the participants agreed to take part in the experiment they signed an informed consent form.

3. Before the experiment started, the participants filled in a demographics questionnaire.

4. The physiological sensors measuring Ecg, Electromyography, Gsr and Electroencephalography were attached, even though all of these physiological measurements were not used in the analysis. The participants were asked to relax for four minutes in order to acquire a baseline recording of physiological data.

5. Each participant performed the experiment with the four conditions: *Solo*, *Human Collaborator*, *Direct Robot Collaborator* with the direct gestured movements and *Non–Direct Robot Collaborator* with the Non–Direct gestured movements. The order of the four conditions was counterbalanced between the participants in order to minimize the ordering effects and explore the difference in elicited physiological affect without the sequential effects. Therefore, sequential learning of the task condition was minimized between the subsequent conditions. Each experimental condition was conducted as follows:

    (a) A participant played a ToH game.

    (b) After a trial was finished, a pause of five minutes was allowed between the trial runs, where the questionnaire data were collected.

    (c) The operator instructed a participant what game task to perform next through the laptop placed next to the participant. A participant was allowed a rest of one minute before starting with the task trial. The operator controlled the laptop using a remote desktop.

The participants were allowed for a six minutes of rest in total between the trials. As performed in the study from (Burgoon and Bonito, 2002), the four conditions (*Solo*, *Human Collaborator*, *Direct Robot Collaborator* and *Non–Direct Robot Collaborator*) were presented to each participant. For each of the four conditions a participant repeated each condition three times one after the other (thus in all, a total of 12 ToH games were played per participant). Each experimental session took around 90 minutes to complete.

10.4.4 Data Collection

The physiological signals were acquired using Biosemi Active Two[1] physiological data acquisition system and its accompanying ActiView 9 software. ECG was measured at the chest using two 16–mm Ag/AgCl spot electrodes in a three-lead unipolar modified chest configuration: the two active electrodes were placed on the right collar bone and the lowest rib on the left side, and the ground electrode was placed on the left earlobe. The signal was amplified, band–pass filtered at 10–40 Hz, and 16–bit digitized; furthermore, it was analyzed by a program that detects R–spikes in ECG, and calculates consecutive R–R intervals. ECG measures the rate and regularity of heartbeats by detecting the peaks of highly positive R–waves in the signal (Houghton and Gray, 2014). The beat–by–beat values were edited for outliers from artifacts or ectopic myocardial activity and linearly interpolated. GSR was measured using surface electrodes attached to the palmar surface of the middle phalanges from the middle finger and the index finger of the non-dominant hand (to reduce mechanical pressure susceptibility). This area, together with the palm of the hand and soles of the feet, is where eccrine sweat glands are most densely distributed, providing the strongest signal variations. Before the electrode application, the participants washed their hands with water and a soap. The temperature and humidity were held constant across the sessions because of GSR susceptibility to the influences of temperature and humidity at the points where the sensors were attached. The sampling rate was fixed at 2048 Hz for all channels. Appropriate amplification and band–pass filtering were performed and subsequently down-sampled to 256 Hz upon data reduction.

10.4.5 Data Reduction and Analysis

The data reduction was performed using Ledalab software for GSR (Benedek and Kaernbach, 2010). Furthermore, Kubios software (Tarvainen et al., 2014) and the HRV Toolkit[2] were used for ECG. This data were compared across the condition differences (*Solo*, *Human Collaborator*, *Direct Robot Collaborator* and *Non–Direct Robot Collaborator*) and the individual differences for the same trials (comparing the responses across individual moves for each condition).

GSR was measured in microsiemens (muS) and analyzed offline. GSR includes short–term phasic responses to specific stimuli, and relatively stable longer–term tonic levels (Andreassi, 2007). In continuous stimulus settings, the most common measures of GSR are Skin Conductance Level (SCL) and Skin Conductance

[1] `http://www.biosemi.com`, accessed 10/12/2018 09:16
[2] `http://physionet.org/tutorials/hrv-toolkit`, accessed 10/12/2018 11:36

Response (Scr), where their changes are thought to reflect general changes in autonomic arousal (Braithwaite and Watson, 2013). The authors stated that the Scr signal is suitable for assessing the intensity of single (phasic) emotions, but changes in the overall (tonic) level are rather inert, thus valid for the trials longer than two minutes, such as the overall session in this experiment. Changes in arousal within periods shorter than two minutes are not likely indicated using the Scl. This problem is particularly limiting in trials shorter than two minutes. When the Scl temporal precision is insufficient, the rapidly reacting phasic changes (Ns–Scr) seem to indicate a more promising focus: their number during a given time period is a prominent phasic–based indicator of arousal (Boucsein, 2012), such as a collaborator condition trial in this experiment.

The raw Gsr signal was down–sampled by a factor of eight (from 2048 Hz to 256 Hz) to remove the high frequency measurement noise and then smoothed by a 25 ms moving average window. The continuous decomposition analysis was performed. The phasic skin conductance detection algorithm used the following heuristics for the valid peak identification for a particular Scr: the slope of the rise to the peak should have exceeded 0.05 muS/min; the amplitude should have exceeded 0.05 muS; and the rise–to–peak time should have exceeded 0.25 s. Standard threshold was set to a minimum amplitude of 0.05 muS (Boucsein, 2012; Dawson et al., 2007), because any function of Ns-Scr and amplitude remains highly sensitive to the threshold that distinguishes it from noise. Such noise may be a consequence of the instruments' quality, the environment, and/or interpersonal differences. Once the phasic responses were identified, the rate of responses was determined. All the signal points that were not included in the response constituted the tonic part of the Scl signal. The slope of the tonic activity was obtained using a linear regression. Another feature derived from the tonic response was the mean tonic amplitude.

Heart rate variability (Hrv) is highly correlated with emotions (Appelhans and Luecken, 2006; Lane et al., 2009). Two measures of Hrv are the standard deviation of normal-to-normal heartbeat intervals in the time domain (Sdnn) and the ratio of low and high frequency powers (Lf/Hf) (Acharya et al., 2006). Wang and Huang (Wang and Huang, 2014) stated that Sdnn and Lf/Hf were employed as two dimensions in the physiological valence/arousal model, where evidence revealed that Sdnn was a good physiological indicator of valence (Kemp et al., 2011; Geisler et al., 2010). The total variance of Hrv increased with the length of analyzed recordings (Saul and Albrecht, 1987). Thus, it is not advisable to compare the Sdnn measures obtained from the recordings of different durations. However, the durations of recordings used to determine the Sdnn values (and similarly the other Hrv measures) were standardized to a minimum of 5 min recordings for the short-term. Generally, the Sdnn levels for the participants with positive affect were found to be higher than for negative one (Geisler et al., 2010). For even shorter recordings, main spectral components of the LF/HF ratio were distinguished in a spectrum calculated for the short-term recordings from 2 to 5 min (Akselrod et al., 1981; Hirsch and Bishop, 1981; Malliani et al., 1991; Pagani et al., 1986; Sayers, 1973).

The raw ECG signal was filtered using a high pass filter with a cutoff frequency of 0.1 Hz. The R-peak detection algorithm performed band-pass filtering of the raw ECG signal and the signal was then smoothed by a 10 ms moving average window. The peaks were then detected in the resulting signal and the detection heuristic rules were applied to avoid missing R peaks or detecting multiple peaks for a single heartbeat. These rules included obtaining the amplitude threshold (the difference between a peak and the corresponding inflection point) at which a peak is considered a beat: enforcing a minimum interval of 300 ms and maximum interval of 1500 ms between the peaks; checking for both positive and negative slopes in a peak to ensure that the baseline drift is not misclassified as a peak; and backtracking with the reexamination/interpolation when a missing peak was detected. Generally, the average change of a heart rate is expected to range between 2–15 bpm (Bradley and Lang, 2000). The chosen interval threshold between the peaks was well above the rate of heart rate change due to the genuine heart acceleration. Data were visually inspected for the artifacts which were subsequently corrected. The R-R intervals were extracted. The time-domain features of Inter Beat Interval, such as the mean and standard deviation, were computed from the detected R peaks. Inter Beat Interval variability was explored by performing a power spectral analysis on the data using Inter Beat Interval to pinpoint the sympathetic and parasympathetic nervous system activation related with the different frequency bands. Ectopic beats, arrhythmic events, missing data and noise effects could have altered the estimation of HRV, therefore the proper interpolation (or linear regression or similar algorithms) on the preceding/successive beats on the HRV signal or on its auto-correlation function could have reduced this error (Kamath and Fallen, 1995). The previously described interpolation steps were performed in Kubios software, while the variables of Heart rate, HRV, SDNN and LF/HF ratio were extracted in the HRV Toolkit.

10.4.6 Hardware and Software Systems

The hardware system contained two Adept Viper S650[3] robot arms with Robotiq Adaptive 2-finger Grippers[4] as the end effectors. The Microsoft Kinect camera

[3] `http://www.adept.com/products/robots/6-axis/viper-s650/general`, accessed 10/07/2017 09:09

[4] `http://robotiq.com/en/products/industrial-robot-gripper`, accessed 10/12/2018 09:11

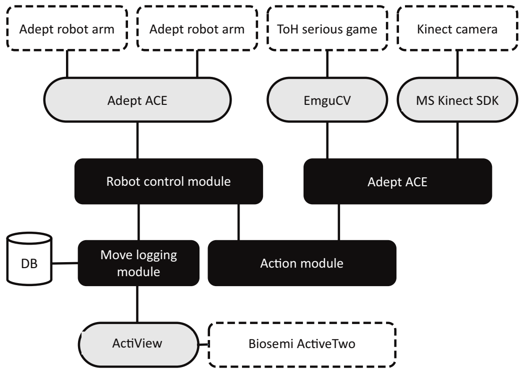

Figure 10.3: The overview of the system for the ToH serious game. The gray boxes were third-party modules for communication with the hardware, while the black boxes were controller modules for the hardware and the serious game, and the white boxes represent the hardware components. The experiment data for analysis were logged to the database (DB).

was used to track the moves made by the humans and robots during the ToH game. A camera was also used for surveillance of the participants and the robot arms, in case of an emergency. A single PC running Windows 7 was controlling the system.

An overview of the software system is shown in Figure 10.3. The Action module was the core of the software system. The Scene module provided information about the moves made and which player is next to make a move. All moves and their corresponding timestamps were stored on the disk using the MoveLogging module. The ActiView 9 was the data recording software for the Biosemi ActiveTwo system, with the aforementioned sensors attached. The RobotControl module was responsible for executing a move. The RobotControl module used the Adept ACE software to control the robot arms. It was used for the movement of robot arms to pick up and drop the specified game disk, and when to close/release the grippers. The Scene module used Microsoft Kinect SDK[5] to connect to the Kinect camera, and the EmguCV5 (a C# version of OpenCV[6]) library to construct a scene of the current game state, as to monitor the game state by following the moves made by humans and robots.

[5] `https://developer.microsoft.com/en-us/windows/kinect`, accessed 10/12/2018 09:14
[6] `http://opencv.org`, accessed 10/12/2018 09:14

10.5 Results

Seventy participants performed all of the experimental conditions (*Solo*, *Human Collaborator*, *Direct Robot Collaborator* and *Non–Direct Robot Collaborator*) of solving the ToH, as they were all executed in a counterbalanced fashion. Physiological arousal was computed from Gsr data, and physiological valence was computed from Ecg data. The differences were analyzed using one-way and two-way Anova. The correlations were explored using a Pearson product-moment correlation index. The analysis was performed using the Spss software with the alpha level at 0.05. Prior to the analysis, Ecg and Gsr data used in the analysis were normalized. The data showed no violation of normality, linearity or homoscedasticity (outliers were removed before the analysis using z-scores for standardized values of ± 3.0 or greater). The participants reported previous experience with robotics on a seven-point Likert scale, where 1 meant "no experience" and 7 meant "familiar experience" ($\mu = 1.7$ $\sigma = 1.095$ $N = 70$). The differences between the participants were analyzed based on the reported values and the experienced outliers were identified. Six participants from the experienced outliers group were removed from the analysis to exclude the effects of participants' familiarity on the experience with the robot collaborators, which resulted in the 64 valid data samples. Moreover, 43 of 70 participants have not had any previous experience with the ToH.

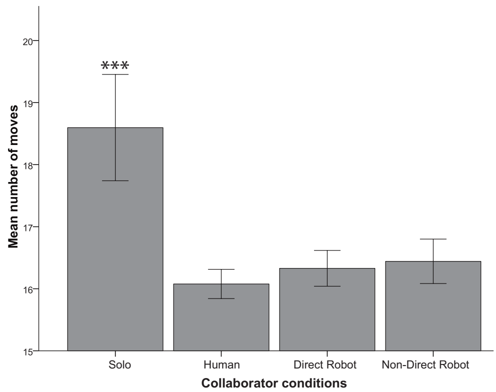

Figure 10.4: Average number of moves per trial in the ToH serious game for each collaborator condition with 95% confidence interval. Stars (***) indicate a significant difference between the Solo condition in contrast to any of the collaborator conditions, at the p <.001 probability level.

10.5.1 Collaborator Conditions and Performance

Regarding the investigation to understand the impact of the collaborator conditions on the performance on a game task (H1), all collaborator conditions were reported as comparable in performance. The analyses examined the performance in regard to the number of moves made by reaching the final configuration of disks in the ToH serious game. Both *Robot Collaborator* condition groups were merged into a single data sample and compared against the *Human Collaborator*. There was no statistically significant difference in the total number of moves between the human and robot collaborator condition groups, as determined by one-way Anova (F(1,569) = 3.705, p > .05), shown in Figure 10.4, where a higher value reflects worse performance. However, a higher number of moves in the *Solo* condition in contrast to any of the collaborator conditions was expected since the participants were not expected to know the most optimal solution for the ToH game task. The difference was statistically significant between the groups as determined by one-way Anova (F(3,751) = 20.807, p < .001). This proof does not support the H1. These results showed that there was no significant difference in the performance between a human and a robot collaborator condition. As expected, the *Solo* condition had a higher number of moves made than any of the collaborator conditions.

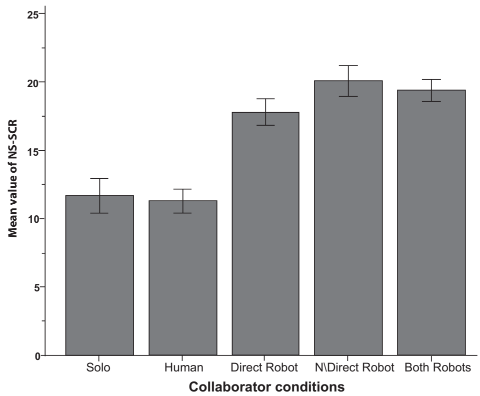

Figure 10.5: The average physiological arousal values measured by NS-SCR variable in the TOH serious game for each collaborator condition with 95% confidence interval. A significant difference ($p < .001$) is observable for both robot collaborators in comparison to the *Solo* and *Human Collaborator* conditions, where the physiological arousal indicator NS-SCR was statistically significantly higher ($p = .017$) for the *Non–Direct Robot Collaborator* compared to the *Direct Robot Collaborator* condition.

10.5.2 Physiological Arousal

In order to examine the hypotheses concerning the elicited physiological arousal for the collaborator (H2a and H2b) conditions, a series of analyses were performed to understand its impact on the performance on game task. The serious game used in this study was found to elicit the relative physiological arousal value (measured with SCL) of 765.209 muS ($\sigma = 835.545$ muS) where the overall arousal value was normalized against the baseline. Therefore, this result suggested that the serious game elicited a high physiological arousal overall. Furthermore, there was a statistically significant difference between the collaborator conditions as determined by one-way ANOVA ($F(3{,}744 = 58.881$, $p < .001$). A Tukey post-hoc test revealed that the physiological arousal indicator NS-SCR was statistically significantly higher for the *Non–Direct Robot Collaborator* (19.59 ± 8.05, $p = .03$) compared to the *Direct Robot Collaborator* (17.43 ± 6.76) condition. Both robot collaborator conditions were statistically significantly higher (18.51 ± 7.49, $p < .001$) than the *Solo* condition (11.62 ± 9.11) and the *Human Collaborator* condition (11.01 ± 6.08), as shown in Figure 10.5. From these results the participants seemed to elicit a higher physiological arousal in the *Non–Direct Robot Collaborator* condition. These findings lend support for the H2a, where the robot collaborators elicited a higher physiological arousal than the human collaborator. Moreover, the *Non–Direct Robot Collaborator* elicited a higher arousal than the *Direct* one, but it made no influence on the performance.

The worse performing participants reported higher arousal values after each round, as the Pearson product-moment correlation was run to determine the relationship between the participant's number of moves and their physiological arousal indicator Ns-Scr. There was a strong, positive correlation between the number of moves per round and the physiological arousal indicator Ns-Scr values, which was statistically significant ($r = .179$, $N = 739$, $p < .001$). These results provided partial support for the H2b and indicated that there was a difference within the elicited physiological arousal that showed an effect on the performance on the game task.

Furthermore, two-way Anova was performed to examine the effects of the collaborator conditions and the physiological arousal on the number of moves (performance) on the game task. The participants were grouped based on lower and higher physiological arousal elicited. There was a significant interaction between the effects of the collaborator conditions and physiological arousal on the number of moves ($F(2, 556) = 8.902$, $p < .001$). The simple main effects analysis showed that the collaborator conditions significantly affected the performance when physiological arousal was lower ($p < .001$), with better performance associated with both robot collaborators compared to the *Human Collaborator*. Between the robot collaborators, better performance was associated with the *Non–Direct Robot Collaborator*, after which comes the *Direct Robot Collaborator* one ($p < .001$).

10.5.3 Physiological Valence

In order to examine the hypotheses concerning elicited physiological valence for the robot collaborator conditions (H3a and H3b), a series of analyses were conducted to understand its impact on the performance on the serious game task. The serious game task environment presented in this study was found to elicit the relative physiological valence score (measured with Sdnn) of .624 s ($\sigma = 1.07$ s), where the overall physiological valence value was normalized against the baseline. Therefore, this result suggested that the serious game elicited a high (positive) physiological valence overall. Furthermore, there was a statistically significant difference between the collaborator conditions as determined by one-way Anova ($F(3{,}732) = 3.575$, $p = .014$). A Tukey post-hoc test revealed that the valence indicator Lf/Hf was statistically significantly higher for the *Non–Direct Robot Collaborator* (3.081 ± 1.869, $p = .008$) compared to the *Human Collaborator*

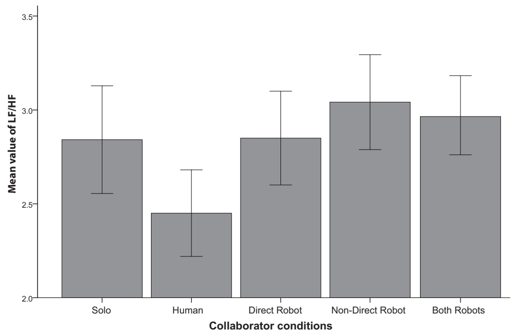

Figure 10.6: The valence indicator LF/HF was statistically significantly higher ($p = .008$) for the *Non–Direct Robot Collaborator* compared to the *Human Collaborator* condition. While both robot collaborator conditions were statistically significantly higher ($p < .001$) than the *Solo* condition and the *Human Collaborator*, there was no statistically significant difference ($p > .05$) between the *Solo* and the *Direct Robot Collaborator* condition.

(2.453 $\pm$ 1.724). Both robot collaborator conditions were statistically significantly higher (2.967 $\pm$ 1.841, $p = .007$) than the *Human Collaborator* condition (2.453 $\pm$ 1.724), as shown in Figure 10.6. There were no statistically significant differences between the *Solo* and the *Direct Robot Collaborator* ($p > .05$), as shown in Figure 10.6. From these results the participants seemed to elicit a higher physiological valence for the *Non–Direct Robot Collaborator*, compared to the *Human Collaborator* condition. These findings lend support for the H3a, where the *Direct Robot Collaborator* elicited a comparable physiological valence as the *Human Collaborator*.

Overall in the experiment, the participants were found to perform the task equally well regardless of the physiological valence found, as no significant correlation was found between the valence indicator LF/HF and the number of moves ($p > .05$). These results provided support for the H3b and indicated that while there were differences across the collaborator conditions regarding the elicited physiological valence, this had no effect on the performance.

10.5.4 Limitation

The limitations of this study were several. The collaborative serious game context of this experiment might have limited the generalizability towards any collaborative task context between humans and robots. Future studies would need to investigate further into the other task contexts. The sample population of the participants was entirely made of the college undergraduates and graduates. As social robots will be integrated in various contexts, future research should deepen the understanding of different populations interacting with the robot collaborators. Furthermore, there might have been fatigue effects present in this experiment, but it was assumed that they were minimal. This statement was supported by the results showing an increased performance during the later trials of the game task for each participant. The experiment lasted for about 90 minutes, with the time on the task around 30 minutes, which was comparable to other similar experiments reported.

10.6 Discussion

The intention of this research is to investigate the performance of participants on the serious game task in a proximate interaction with the human and robot collaborators eliciting physiological affect. Furthermore, this study investigated the effects of elicited physiological arousal and valence on the human-robot collaborative performance on the game task. This investigation was based on the social cues of the collaborating robots (i.e., gestured movement and speed). Overall, collaborators in the ToH serious game elicited physiologically arousing, pleasantly valenced experience. The participants performed equally well on the game task, regardless of the collaborator conditions. Furthermore, the robot collaborators elicited a higher physiological arousal than the human one, while the *Non–Direct Robot Collaborator* condition elicited a higher physiological arousal than the other collaborator conditions. The participants with a higher elicited physiological arousal performed worse on the game task. The *Non–Direct Robot Collaborator* condition elicited a higher physiological valence, compared to the *Human Collaborator* one. Results further indicated that the physiological valence had no statistically significant effect on the performance on the game task.

Considering the H1, as the difference in the participants' performance was not statistically significant between any of the collaborator conditions, indicating that the collaboration with robot partners might be as effective as collaboration with human ones. It is possible that the participants may have been highly focused on the game task since the collaboration with a physical entity eliciting diverse behaviors and strong emotional responses might have promoted a higher focus on the task (Hocine and Gouaich, 2013; Chen, 2007; Csikszentmihalyi, 2008). As the worst performance was found in the *Solo* condition, it is possible that the participants left to their own skills had more room for the non-optimal moves, since the help of the collaborators was not available.

The relevance of the previous claims is further supported through the H2a and the H2b, exploring the performance in regard to the elicited physiological arousal on the collaborative serious game task. The robot collaborators elicited a higher physiological arousal than the human ones, while the *Non–Direct Robot Collaborator* elicited a higher physiological arousal when compared to the *Human Collaborator*. The results indicate that people are sensitive to the robots' social cues regarding physiological arousal in the context of collaborative serious games, supported by the previous investigations on social cues in the context of HRI (Butler and Agah, 2001; Goodrich and Schultz, 2007). As a higher physiological arousal results in a lower performance, this may indicate that high physiological arousal is associated with worse performance in the context of collaborative serious games, supported by the previous investigations on connection between arousal and performance on a task (Yerkes and Dodson, 1908; Csikszentmihalyi, 2008). The authors state that the performance is positively correlated with physiological arousal up to the point when the level of arousal becomes too high and the performance decreases. Possibly there was a high physiological arousal elicited in this study, especially if we consider the "lower" physiological arousal group which was the only one that showed the statistically significant effect of collaborator conditions on the performance. The other "higher" physiological arousal group had no statistically significant effect on the performance. These findings indicate that serious games might elicit high physiological arousal which may have disrupting effects on the performance on the game task.

Regarding the H3a and the H3b, investigating on the performance in regard to the elicited physiological valence on a collaborative serious game task. The *Non–Direct Robot Collaborator* elicited higher (positive) physiological valence compared to the *Human Collaborator*, indicating as well that people are sensitive to robots' social cues regarding physiological valence in the context of collaborative serious games. As the participants performed equally well regardless of the elicited physiological valence, this may suggest that physiological arousal has a more profound effect on the performance than physiological valence in the context of collaborative serious games, as supported by the previous studies exploring characteristics of robot behavior in HRI (Butler and Agah, 2001; Goodrich and Schultz, 2007).

10.7 Conclusion

The experiment was conducted in a realistic setting free of the experimenter presence, in which the participants had actually interacted with the autonomous robot collaborator, in contrast to the Wizard-of-Oz experiments (Sisbot and Alami, 2005). The contributions of the present study include the advances in both theoretical and practical understanding of physiological affect in the context of HRI. Moreover, they include the design of a serious game with the robot collaborators that could elicit physiological affect essential to HRI in such a context. This is particularly important if one considers that there is a direct link between the embodiment and information provided by the social cues of robot collaborators eliciting physiological affect, which is dependent on a physical interaction and a serious game environment itself (Pfeifer et al., 2012). This research outlines a method for the objective and continuous measurement of the physiological affect in collaborative HRI, in the context of collaborative serious games. Moreover, it supports the notion that even non-humanoid robot collaborators can display social cues and elicit physiological affect in their human partners.

Overall, the collaborators in this study created a physiologically arousing, high (positive) physiologically valenced serious game environment. As a number of moves per trial in the serious game was consistent across all the collaborator conditions studied, these findings indicate that the participants' performance on the serious game task is comparable between the human and robot collaborator conditions. Regarding autonomous robots, this study found evidence of higher physiological affect elicited (arousal and valence) in contrast to their human collaborator counterparts, while still indicated a comparable performance on the game task between them. These findings motivate the introduction of autonomous robots as partners in the context of collaborative serious games, where the same performance benefits may be achieved as with using human ones. The *Non–Direct Robot Collaborator* condition elicited a higher physiological arousal and a (positive) valence, compared to the *Human Collaborator*. Moreover, it elicited a higher physiological arousal than the *Direct Robot Collaborator* condition, indicating that the careful design of *Direct Robot* partners can leverage different social cues to elicit target physiological arousal in the context of collaborative serious games. Furthermore, such context my witness a more positive valence elicited when using *Non–Direct Robot* partners instead of human ones. This may be important as robots get introduced to different aspects of human lives, possibly as team members (Fiore et al., 2013). The current study supports the notion that understanding physiological affect underlying such collaborative HRI from the human perspective, it would be possible to design more personalized serious games with intelligent robots which act together with human partners eliciting relevant physiological affect (Xin and Sharlin, 2006). This may contribute to improving the quality of HRI informing the design of such collaborative serious games. On the other hand, one has to be careful when designing serious games which elicit high physiological arousal, as such high levels of physiological arousal may be correlated with lower performance (Csikszentmihalyi, 2008; Yerkes and Dodson, 1908). In contrast, physiological valence may not have such a significant effect.

If one considers designing serious game environments that elicit lower physiological arousal using robot instead of human collaborators, than one might witness an increase in the performance on a serious game task. As this study found evidence that the better performance was associated with the robot collaborators compared to the human ones, only for the “lower” physiological arousal group which was the only one showing the statistically significant effect of the collaborator conditions on the performance. Finally, if one considers that robots possess the physical-virtual duality and have access to a game task information, one can clearly see why a choice of robot collaborators in serious games would be a sound choice.

Taking a step forward with using the physiological measurements, recognition of affective states is expected by cooperating humans, which may allow them to be aware of their emotions through the presentation of a sufficient feedback (Picard, 2000). Future studies should investigate the recognition of participants’ emotions online using physiological measurements to adapt the robots’ behavior in a closed-loop social interaction (Liu et al., 2008), as the embodiment is a powerful concept in the development of the adaptive autonomous systems (Pfeifer et al., 2014). Taking these findings together with the proposition from Breazeal (2002) that these investigations would seek to improve the quality of Hri, this study is an attempt to inform the design of more natural, intuitive and familiar serious games and robot collaborators.

Chapter 11

Practicing Emotion–Regulation Through Biofeedback on the Decision–Making Performance in the Context of Serious Games: A Systematic Review

Abstract

Evidence shows that emotions critically influence human decision–making. Therefore, emotion–regulation using biofeedback has been extensively investigated. Nevertheless, serious games have emerged as a valuable tool for such investigations set in the decision–making context. This review sets out to investigate the scientific evidence regarding the effects of practicing emotion–regulation through biofeedback on the decision–making performance in the context of serious games. A systematic search of five electronic databases for the systematic search (Sco-

pus, Web of Science, IEEE, PubMed Central, Science Direct), followed by the author and snowballing investigation, was conducted from a publication's year of inception to October 2018. The search identified 16 randomized controlled experiment/quasi–experiment studies that quantitatively assessed the performance on decision–making tasks in serious games, involving students, military, and brain–injured participants. It was found that the participants who raised awareness of emotions and increased the skill of emotion–regulation were able to successfully regulate their arousal, which resulted in better decision performance, reaction time, and attention scores on the decision–making tasks. It is suggested that serious games provide an effective platform validated through the evaluative and playtesting studies, that supports the acquisition of the emotion–regulation skill through the direct (visual) and indirect (gameplay) biofeedback presentation on decision–making tasks.

11.1 Introduction

Biofeedback allows individuals to visualize specific physiological parameters to raise awareness and improve *emotion–regulation* (Er) so that the aversive behavior may be changed (Repetto et al., 2013). Unfortunately, these physiology visualizations are nonintuitive to many individuals. Compared to traditional approaches, *serious games* are interactive media which can provide a context for the biofeedback presentation and on–line perception of emotional responses, which are reflected on the changes in the individuals' physiology (Yannakakis et al., 2016). Such biofeedback methods provide a shortcut to raising awareness of internal bodily states and emotions, which have been found to critically influence human decision–making (Adam and Gamer, 2011; Loewenstein, 2000). In the context of serious games, physiological sensors are used to monitor players' emotional states during gameplay, while the game mechanics are then adapted to reward the target behaviors (Parnandi et al., 2013). Previous studies have shown that such 'serious game biofeedback' approach facilitates skill–acquisition and transfer (Parnandi and Gutierrez-Osuna, 2018). Emotions and decision–making biases are even beyond conscious awareness, and it is therefore highly interesting to regulate such emotional dispositions (Fenton-O'Creevy et al., 2011). Therefore, awareness of emotions is a necessary prerequisite to be equipped with the skills to decide whether it is better to follow an emotional response or not (Gross, 1998b).

Emotion–regulation can help to mitigate emotion–related decision biases and eventually lead to a better decision performance. Emotion–regulation using biofeedback has been extensively investigated. Nevertheless, serious games have emerged as a valuable tool for such investigations set in the decision–making context (Yannakakis et al., 2016). This interactive approach to such an investigation is highly context–dependent, where Er skill–acquisition and transfer are 'maximized' when individuals are provided with engaging context–dependent serious games, which have emerged as a tool that fits this purpose (Sliwinski et al., 2017). Furthermore, they can parallel the qualities of the traditional approaches regarding beneficial effects on attention and Er, while providing valuable conditions to improve accessibility, for example by giving clear instructions and multisensory feedback (Sliwinski et al., 2017).

Previously published reviews argue that the future investigation would benefit from a systematic investigation of experimental work (Miller, 2015), examining in detail which game features are most effective in promoting engagement and supporting skill–acquisition/transfer (Culbert, 2017; Boyle et al., 2016). The investigated reviews argue that the research on serious games is already an established field of science, where decision–making questions are a burning issue.

These reviews were exclusively concerned with individuals' affective experience and motivation in serious games (Carter et al., 2014). Furthermore, previous reviews provide a useful overview of games research, but the escalating volume of such research suggests that it would be more useful for future literature reviews to focus on more specific issues (Boyle et al., 2016), which we tend to follow in our review.

During compiling this manuscript, a relevant literature review was published in the field of Er and serious games (Villani et al., 2018). In contrast to the presented work, the mentioned review focuses on the investigations regarding well–being, relaxation, and Er training in clinical interventions. It is structured around the overall amount and quality of serious games regarding their affective and modality focuses. Since the selected studies do not overlap between these two reviews, this gives evidence that each of them is focusing on a different aspect of Er is serious games.

We have conducted a systematic, structured literature review to identify and investigate all the reported studies and summarized the current state–of–the–art in the effects of practicing Er through biofeedback on the decision–making performance in the context of serious games. We named this context with the acronym Beds further in the text. Surveying the previous reviews found that there was no structured literature review in the mentioned area.

11.2 Related Work

The various overlapping concepts within the models of emotions and affect are translated through psychology to computer science, especially its determinants, measures and how these were related to Er. These various concepts make it difficult to develop valid measures and a common understanding of emotions and affect, especially in the context of decision–making in serious games. According to Russell, emotions can be classified by the independent components of arousal and valence Russell (1980). According to this circumplex model defined by these two components, arousal represents the excitement level whereas valence defines whether the current emotional state is positive or negative. Physiology allows us to extract and interpret both valence and arousal in relation to emotional stimuli (Leng et al., 2007; Anttonen and Surakka, 2005; Cacioppo et al., 1986; Winton et al., 1984). Apart from the circumplex model, a discrete model of emotions takes into account the level of presence of basic emotions like happiness, sadness, engagement, anger, fear (Bontchev and Vassileva, 2016; Changchun et al., 2009). While emotions enable us to take advantageous decisions (Gross, 2009;

Seo and Barrett, 2007), there is a possibility that emotions also hinder advantageous decision–making, especially if they are not accurately processed and get ‘out of control’ (Adam et al., 2013; Bechara and Damasio, 2005; Shiv et al., 2005; Loewenstein, 1996). Individuals can lose control over their actions under the influence of high arousal, which can significantly threaten rational cognition and decision–making (Astor et al., 2013b; Kuhnen and Knutson, 2005). Therefore, a need arose to raise awareness of emotional processes and how they can affect decision–making performance.

11.2.1 Decision–Making

Currently, psychology sees decision–making as a dual–process framework which works in two fundamentally different ways depending on the context: the slow and analytical rational system; and the fast and intuitive experiential system (Kahneman and Frederick, 2002). Decision–making may require a delicate balance between the activation of one framework mechanism or the other, which may lead to reduced performance (Kuhnen and Knutson, 2005). The affect heuristic states that human decision makers integrate context–specific affective feelings into perception and their decision–making process. Therefore the context of a task has been identified as an important factor of how affective stimuli are processed (Rolls and Grabenhorst, 2008).

The somatic marker hypothesis concluded that emotional responses to information events seem to be both a source of biases, as well as an important mechanism for advantageous decision–making, where somatic markers beneficially guide our focus of attention (Bechara and Damasio, 2005). According to the theory, decisions are aided by emotions in the form of elicited bodily states during the deliberation of future consequences, which makes different options for advantageous or disadvantageous behavior (Naqvi et al., 2006). Arousal, whether caused by positive or negative valence, may be equally (and some argue even more) important than valence for decision–making performance. It has been found that individuals aware of their emotional states avoid being overpowered by their emotions. These individuals need to develop skills to interpret and down–regulate those affective states of high arousal (Fenton-O’Creevy et al., 2012a; Peterson, 2007), which may result in enhanced decision–making performance (Biais et al., 2005).

11.2.2 Emotion–Regulation

Evidence shows that Er contributes to an advantageous decision–making (Fenton-O'Creevy et al., 2011), where individuals with lower Er skills are prone to make disadvantageous decisions with undesirable consequences (Sütterlin et al., 2011). The evidence further shows a significant relationship among the effective Er, increased awareness of emotional states and a sophisticated skill in down–regulating high arousal (Fenton-O'Creevy et al., 2012b). As Er can be both consciously and unconsciously processed (Williams and Bargh, 2009), increased emotional awareness is crucial for improving it (Culbert, 2017). Increasing awareness of emotional states can improve the conscious awareness of one's physiological processes – interoception (Critchley et al., 2004). This review argues that using Er in the Beds context, one can practice the skill of down–regulating high stress and arousal to perform better on a decision–making task.

People tend to use one of the two broad Er strategies to deal with emerging emotions (i.e., cognitive reappraisal and suppression) while facing difficult and stressful tasks (Wallace et al., 2009). The authors state that while response–focused Er strategy is 'suppression,' where it aims at altering and controlling the experiential, behavioral, and physiological response when the emotion has already unfolded completely; antecedent 'cognitive reappraisal' Er strategy applies while the emotion has still not reached its peak while unfolding. Suppressors tend to constantly push down emotions, ignoring the fact that they exist and are continuously affecting them, where the degree of inhibitory control is negatively correlated with susceptibility to framing effects in a decision task (Sütterlin et al., 2011). The authors state that both Er strategies take up cognitive resources, but suppressing emotions generally takes up more cognitive resources in comparison to the reappraisal strategy when encountering undesired emotions. Hence, emotional suppression can eventually take up so many cognitive resources that it can reduce one's performance in decision–making tasks compared to the strategy of emotional reappraisal.

Self–regulation is a broader term which includes mental resilience through Er (Algoe and Fredrickson, 2011; Reivich et al., 2011). Emotional resilience (also referred to as mental resilience) is the capacity to withstand, cope or recover from stress and adversity (Rose et al., 2013; Buckley and Anderson, 2006).

11.2.3 Biofeedback

While contingent reinforcement of behavior through biofeedback has been a part of the investigations from the early start, it has not been thoroughly investigated

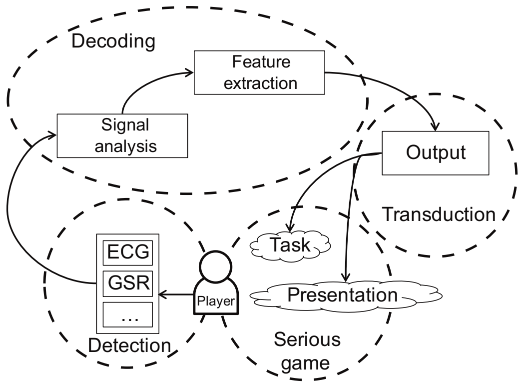

Figure 11.1: The schematic overview of the biofeedback mechanism in serious games.

(Bergman and Johnson, 1972). In contrast to traditional biofeedback, Gilleade et al. (2005) refine this definition of affective feedback by stating that the individuals may not be aware that physiological states are being assessed since the intent is to evaluate natural affective reactions (Kuikkaniemi et al., 2010). The investigation of the effect of incentive upon *heart rate* (Hr) acceleration and deceleration through biofeedback found that subjects in the incentive group (monetary reward) showed better Hr control than those in the non–incentive group (Lang and Twentyman, 1976). *Instrumental conditioning* (Ic) is the process of rewarding and penalizing individuals' behavior based on their decision response, and it is the central mechanism in the acquisition of skills using biofeedback in serious games (Parnandi and Gutierrez-Osuna, 2017), as depicted in Figure 11.1. More specifically, the concept where the target Ic behavior eliminates the occurrence of an aversive stimulus which leads to the reinforcement of that behavior (skill–acquisition), is termed *negative reinforcement instrumental conditioning* (Nr–ic). The somatic hypothesis and research in psychophysiology found that the experience of both emotions and stress are known to be accompanied by a physiological state of arousal, which in turn manifests in changes in physiology (Lehrer and Vaschillo, 2004; Grandey, 2000). Biofeedback in the Nr–ic context forces individuals to down–regulate their arousal levels (i.e., the instrumental response) to reduce game penalty (the aversive outcome) and progress in serious games (Cannon-Bowers and Salas, 1998). Furthermore, Nr–ic increases the likelihood that instrumental behavior will be repeated in the future indicating skill–transfer (Domjan, 2014). Whether the Ic increases or decreases a behavior depends on both the nature of the outcome (i.e., aversive or appetitive) and whether the behavior produces or removes the outcome. Furthermore, reinforcements can be categorized as appetitive (pleasant outcome) and aversive (unpleasant outcome). Therefore, Ic procedures can be classified into four categories (Domjan, 2014): *positive reinforcement* – when the target behavior produces an appetitive outcome, which leads to a reinforcement of that behavior; *punishment* – when the target behavior produces an aversive stimulus, which leads to a reduction in that behavior; *negative reinforcement* – when the target behavior eliminates an aversive stimulus, which leads to a reinforcement of that behavior; and *omission training* – when the target behavior eliminates an appetitive stimulus, which reduces that behavior.

Biofeedback in serious games can be viewed as a form of Ic in which reinforcements (i.e., rewards or penalties in the game) are used to modify voluntary behaviors (e.g., increase or decrease breathing rate). In the Nr–ic context, two types of reinforcement schedules (partial and continuous) were investigated to encourage players to slow down their breathing rates during gameplay, which reduces an aversive stimulus (i.e., random actions in the game) (Parnandi and Gutierrez-Osuna, 2018). In fact, a long history of behavioral research shows that the reinforcement schedule can have a significant impact on skill–transfer, where Nr–ic increases the likelihood that the instrumental behavior will be transferred (Parnandi and Gutierrez-Osuna, 2018).

11.2.4 Physiology

Heart–rate variability (Hrv) is quantified by measuring the interval variability between successive heartbeats (Rr *intervals*) in the *electrocardiogram* (Ecg) signal (Kim et al., 2013). Biofeedback may improve Er and cognitive functioning since evidence shows that Hrv is associated with emotions, as suggested by the somatic hypothesis (Lehrer and Vaschillo, 2004). Previous investigations established a direct connection between the central nervous system and the *autonomic nervous system* (Ans), which is reflected in Hrv (Thayer et al., 2009), mediated through regulation of the vagus nerve (Lane et al., 2009; McCraty et al., 2009).

Evidence shows that large amplitude modulation in Hrv and *respiratory sinus arrhythmia* (Rsa) are associated with certain specific executive functions: attention, flexibility of behavior and control of emotions (Thayer et al., 2009). Rsa refers to the component of the change in Rr intervals that is synchronized to the respiratory cycle (Lehrer et al., 2000). Moreover, Rsa may be a dominant component of the change in the Rr interval when the individual's breathing is at an optimal frequency, which is referred to as 'resonant frequency' or 'coherence.' Maximal regulation occurs at a particular cardiovascular 'resonant frequency' of the baroreflex system correlated to Hrv, typically 0.1 Hertz or a 10–second rhythm (Vaschillo et al., 2006; Lehrer et al., 2000), and it has been found to reflect a balance between the two branches of the Ans, the sympathetic and parasympathetic nervous system. Deep breathing is a method to address the autonomic imbalance that arises from exposure to a stressor (Jerath et al., 2006). It recruits the parasympathetic branch of the nervous system and inhibits the sympathetic action, leading to a regulated state (Parnandi and Gutierrez-Osuna, 2018). Regulating breathing rate towards this resonant frequency shifts the autonomic balance maximizing Hrv and down–regulates high arousal (Jerath et al., 2006; McCaul et al., 1979). Therefore, practicing Rsa through the Hrv biofeedback improves the skill–acquisition of Er by influencing physiology (Moraveji et al., 2011). The goal of Hrv biofeedback is to help individuals increase the relative amount of Rsa in the Hrv signal (Kim et al., 2013). From both a psychological and physiological standpoint Rsa has been shown to be most closely associated with self–regulation (Thayer et al., 2009). Previous investigations found that individuals with greater Hrv had significantly more correct responses in a working memory test and in a 'Continuous Performance Test,' as well as faster reaction times (Hansen et al., 2003). Rsa is also correlated with arousal values inferred from the Hr (Mezzacappa et al., 1998).

Electrodermal activity (Eda) signal reflects changes in the sympathetic nervous system based on changes in the electrical skin conductance in response to external stimuli. Furthermore, Eda activity consists of two main components – tonic and phasic responses. Tonic skin conductance refers to the ongoing or baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event–related changes, caused by a momentary increase in the skin conductance (Kuikkaniemi et al., 2010). Inference of psychophysiological arousal based on Eda measurement has been proven to have a strong content validity (Bontchev and Vassileva, 2016; Fairclough, 2009).

11.2.5 Physiological and Affective Computing

Physiological computing allows for human physiology to be directly monitored and transformed into a control input for any technological system (Fairclough, 2009). Such concept came out of the biofeedback approach that creates a closed–loop feedback design (see Figure 11.1) to teach human participants the necessary skills for autonomic self–regulation. This design enabled the adaptive gaming technologies where the act of self–regulation is integrated directly into the gameplay (Pope et al., 2014). The authors further state that biofeedback systems offer immediate reinforcement effects to facilitate 'training' or 'therapy' as a means to change physiological functioning and improve self–regulation skills. Biofeedback allows an individual to receive on–line feedback on their visceral processes which are usually almost unperceivable (Dawson et al., 2011). On another hand, affective computing draws methods from the original experiments on the psychophysiology of emotion, where psychophysiological changes are used for the detection of emotional categories (e.g., happiness, fear, disgust). Affective computing incorporates dynamic software systems sensing and adapting themselves on–line to the affective states based on psychophysiological measurements of the individual using it (vom Brocke et al., 2012; Picard, 2000). A shift towards more nuanced states that incorporate elements of both emotion and cognition enabled the use of psychophysiological activity from the player, which may be used to inform the response of interactive serious game applications. The potential of physiological computing to tailor each interaction to the specific responses from the individual player provided affective computing with the means to control the adaptive automation and present useful information to the player.

11.2.6 Serious Games

Serious games can be understood as games that aim at purposes other than entertainment alone, and they are engaged in a 'serious' investigation of certain aspects of human endeavors (Liu et al., 2013). Repeatable, highly engaging, and motivating serious games can result in improved training activities and have positive effects on the development of new knowledge and skills (Léger et al., 2011; Anderson and Lawton, 2009). When decision performance may vary based on the individual's affective states, biofeedback–assisted tasks may achieve an increased decision–making performance. This has been supported through the previous investigations which have found that individuals can learn to attenuate the magnitude of their cardiovascular reactions to a variety of behavioral and physical stressors when provided with biofeedback–assisted training (Larkin et al., 1990).

Biofeedback and other traditional methods to practice Er are hard to perceive by individuals, since they would need to invest a significant amount of time in training or need the supervision of a therapist, which can be time and cost prohibitive (Sliwinski et al., 2017). Validated biofeedback self–guided interventions (i.e., meditation, cognitive based therapy, and yoga) suffer from high dropout rates due to the unengaging nature of exercises and lack of motivation (Rose et al., 2013; Davis and Addis, 1999). Furthermore, these techniques teach self–regulation in quiet, controlled settings, which may not generalize to the context of real–world stressors (Driskell and Johnston, 1998). Traditional approaches for practicing Er fail mostly because of the limited capacity for self–monitoring (Astor et al., 2013b). As traditional approaches fail in improving

ER skill–acquisition (Fenton-O'Creevy et al., 2012a), serious games emerged as a new angle introducing technological methods to practicing ER, where meaningful biofeedback information communicates individuals' emotional states on–line (Astor et al., 2013b).

Using biofeedback in serious games offers many advantages to practicing ER. In a stressful situation, the individual's focus is rarely on the physiological and psychological effects of stress. Therefore continuous information about the internal states while playing a highly captivating serious game would offer a possibility for players to become aware and actively use the appropriate coping strategy. Such stressful and highly captivating serious games are tailored to the players' individual skills through biofeedback by reducing the player capacity to play efficiently based on on–line physiological data. Contrary to the unengaging controlled environments which lack motivation and context (Rose et al., 2013; Davis and Addis, 1999; Driskell and Johnston, 1998), serious games can address these shortcomings by providing a context for the stressor and eliciting emotions including stress and arousal (Buckley and Anderson, 2006). Biofeedback skill–acquisition is reinforced with a positive reward through the mechanism of operant conditioning. This is achieved through a provision of physiology information and reinforcement score presented after a successful regulation. Players have to make a decision to administer positive or negative rewards based on their success, where successful regulation is supported by the game score of the decision–making outcome.

11.2.7 Skill–Transfer

For the basic principle of effective ER skill–transfer to work, an environment that induces stress and arousal is required, which may be difficult to recreate in didactic settings or role plays in a briefing room (Bouchard et al., 2012; O'Donohue and Fisher, 2009). The neuroscience perspective might explain the skill–acquisition through using biofeedback in serious games, where performing goal–directed tasks (i.e., playing a serious game) leads to dopamines release in the striatum of the brain (Koepp et al., 1998). The authors found that there is a monotonic increase in striatal dopamine levels released during gameplay (compared to baseline) and that the effect was sustained after the gaming session ended. Such dopamine release is an indicator of memory storage events and attention (Achtman et al., 2008). Furthermore, it is also involved in learning stimuli or actions that predict rewarding or aversive outcomes. Therefore, serious games may facilitate improved awareness of arousal and effective skill–acquisition/transfer of ER (Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017).

11.3 Method

The selection of the publications was made according to an adapted 'The Quality of Reporting of Meta–analyses' (QUOROM) procedure (Bargas-avila and Hornbæk, 2011).

Source selection A preliminary search investigation had suggested that the relevant publications in the BEDS context are spread across multiple scientific journals and conferences. Therefore, a systematic search of five electronic databases (Scopus, Web of Science, IEEE, PubMed Central, Science Direct) was conducted. As there was no systematic review on this subject up till now, a timeframe limitation was excluded; therefore it was set from a publication's year of inception to the October 2018.

Search string In order to identify the relevant papers for this literature review, we used individualized search strategies for the different databases, which included combinations of the following keywords: (game OR games OR gaming OR "serious play" OR "serious playing" OR task) AND (gam* OR play*) AND (emotion* OR affect* AND NOT (affects OR affecting OR affected)) AND (biofeedback OR neurofeedback OR ("neuro–feedback") OR ("neuro–biofeedback") OR neurobiofeedback) AND ((emotion* regulation) OR (modulat* emotion*))). This search identified articles related to serious games with biofeedback for practicing ER mentioned in the body of the paper, including the ones investigating the decision–making on tasks, which were identified later through the selection process. There may be a possibility that we are missing some papers mentioning physiology modalities directly or a specific affect or emotion, but without reference to ER or psychophysiological theories, those might not be relevant for our review. Only articles with *randomized controlled trials* (RCT) or quasi–experiment design with serious games investigating ER and decision–making performance published in refereed peer–reviewed journals and conferences were considered for review. Concepts such as emotions or affect were only included if the authors explicitly equated them with decision–making and ER. Informal literature surveys (e.g., theoretical papers) were excluded, as well as, the studies which did not contain measurable quantitative decision–making performance data and reported results on ER through biofeedback.

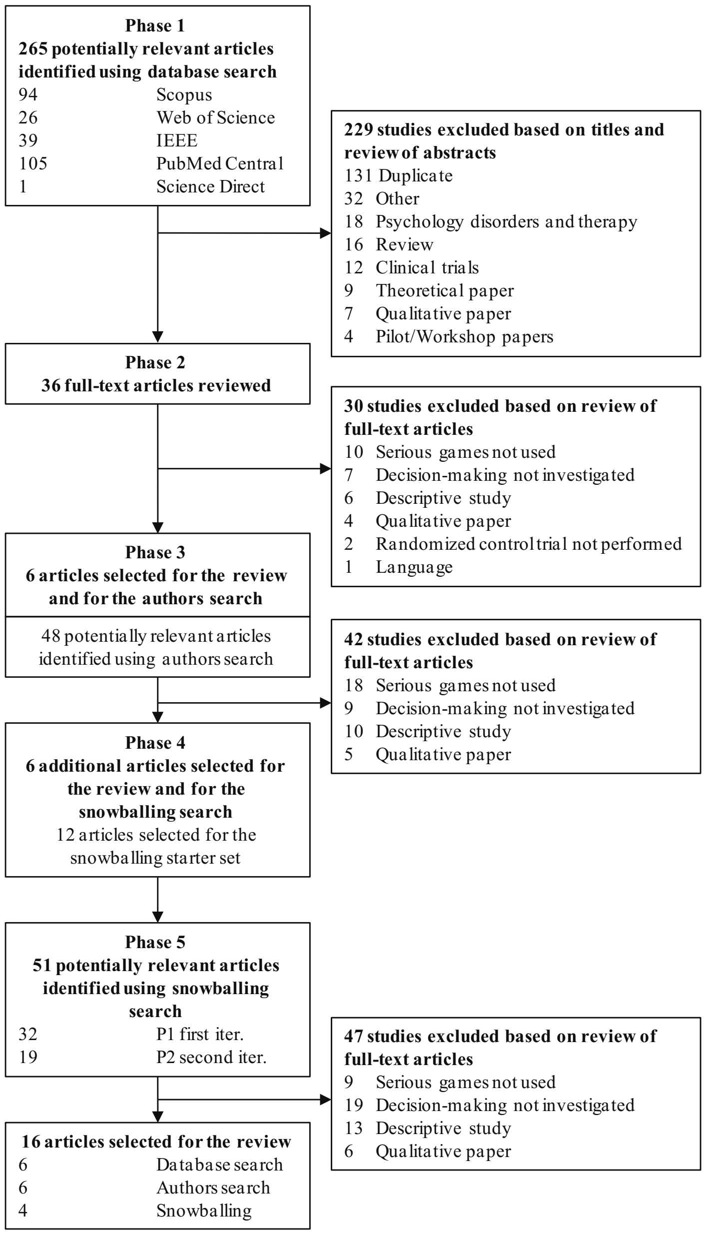

Figure 11.2: The flow of studies throughout the review process

Search procedure Figure 11.2 shows the flow of studies throughout the review process, with the reasons for exclusion detailed at every step. In all five databases, the searches were combined to include all publications. The search resulted in a total of 265 publications. All of the collected abstracts were read and assessed based on their quality, making sure they truly cover the mentioned search terms. Discards were verified by a second rater, narrowing the main body of identified studies down to 36 quantitative studies. Two researchers took part in the identification, coding and quality assessment of the papers. The search was conducted by the first author, whereas both authors were involved in the review process, where Cohen Kappa statistic was derived from the same categorization using 20% of the entries to control for interrater effects. Interrater reliability scored Kappa = .85 ($p < .001$), 95% CI (.788, .067), which is higher than a Kappa value for almost perfect agreement of .8. The papers were scored on the concepts investigated by assigning a value of 0 to 4. Lastly, the publications included for analysis had to be original full papers and written in English. After a preliminary evaluation of the selected abstracts and references by the researchers to determine relevancy, full–length copies of the selected references were secured for the final evaluation. Again, the screening was done by the first author and an independent rater, while potential disagreements were resolved through iterative discussions.

Formulating a suitable search string is a very difficult task, as the terminology used is often not standardized and a large number of irrelevant papers will be found using broad search terms (Wohlin, 2014). This approach may miss important research. Due to the small amount of the initial research papers selected through the database search, the identified authors were checked for the new papers published additionally. Furthermore, backward and forward snowballing was conducted using Google Scholar in addition to the previously mentioned electronic databases, as outlined in the (Wohlin, 2014). Research selected through the database and corresponding authors search has been included as the actual start set for the snowballing approach, and these were included in the systematic literature study.

The numbers of investigated papers through the snowballing method:

- Start set: 12 papers for the start set were included (6 papers selected from the database search and 6 papers selected from the authors search)
- Iteration 1: 32 candidates from backward and forward snowballing from start set and 4 papers were included
- Iteration 2: 19 candidates for inclusion were generated in backward/forward snowballing and no paper was included, and hence the snowballing ends.

Coding of papers Iterative open coding was employed on the 16 selected papers, to identify different categories of research. These categories were then clustered according to the themes and axially coded for qualities such as methodology, results, aims, findings and so forth. Furthermore, the papers selected by the inclusion criteria were analyzed using the data extraction form that was developed by Connolly et al. (2012), to find patterns concerning the main focus of the study.

Quality assurance Papers were quality assessed according to the five criteria (Connolly et al., 2012), where a formal quality score was completed by assigning a point for each criterion assessed: (1) quality of research design using Rct or quasi–experiment design; (2) appropriate methods and analysis; (3) generalizability of findings; (4) the relevance of the focus of the study (including serious games, decision–making, biofeedback, and Er) for addressing the research question and (5) trustworthiness of findings. Studies that scored 1–2 were regarded 'very low' quality studies, while studies that scored 3–4 were classified as 'relevant' quality studies and were included in the review. Due to the contextual nature of biofeedback and Er, the generalizability of findings was jeopardized in all of the studies, rendering 4 as the maximal score through the quality assurance. The body of 36 identified studies from the database search, 48 identified studies from the authors search and 51 identified studies from the snowballing search (135 full–text articles in total) were assessed for quality, out of which 47 studies scored 'very low', 72 studies scored 'non–relevant' and 16 studies selected for this review scored 'relevant'.

Nr.	Authors (Year)	Title	Design	Group Size	Number/Time of Trials
1	Parnandi and Gutierrez–Osuna (2018)	Partial Reinforcement in Game Biofeedback for Relaxation Training	RCT	24	17 min (3 x 5 min + break 2 min)
2	Zafar et al. (2018)	Gaming Away Stress: Using Biofeedback Games to Learn Paced Breathing	RCT	103	8 min
3	Parnandi and Gutierrez–Osuna (2017)	Visual Biofeedback and Game Adaptation in Relaxation Skill Transfer	RCT	24	30 min (6 x 5 min)
4	Zafar et al. (2017)	Playing With and Without Biofeedback	RCT	43	8 min
5	Lara–Alvarez et al. (2017)	Induction of Emotional States in Educational Video Games through a Fuzzy Control System	RCT	20	39 min (2 x 12 min + break 15 min)
6	Astor et al. (2013)	Integrating Biosignals into Information Systems: a NeuroIS Tool for Improving Amotion Regulation	RCT	104	25 min
7	Hilborn et al. (2013)	A Biofeedback Game for Training Arousal Regulation During a Stressful Task: The Space Investor	Quasi—Experimental	5	180 sec
8	Bouchard et al. (2012)	Using Biofeedback while Immersed in a Stressful Videogame Increases the Effectiveness of Stress Management Skills in Soldiers	RCT	60	30 min
9	Jerčić et al. (2012)	A Serious Game Using Physiological Interfaces for Emotion Regulation Training in the Context of Financial Decision–Making	Quasi—Experimental	6	25 min
10	Kuikkaniemi et al. (2010)	The Influence of Implicit and Explicit Biofeedback in First–Person Shooter Games	Within–Subjects	36	2000 sec (10 x 200 sec)
11	Liu et al. (2009)	Dynamic Difficulty Adjustment in Computer Games Through Real–Time Anxiety–Based Affective Feedback	Quasi—Experimental	15	24 min (12 x 2 min)
12	Goodie and Larkin (2006)	Transfer of Heart Rate Feedback Training to Reduce Heart Rate Response to Laboratory Tasks	RCT	14	2 hr
13	Goodie and Larkin (2001)	Changes in Hemodynamic Response to Mental Stress With Heart Rate Feedback Training	RCT	46	35 min (5 x 6 min + break 5 min)
14	Larkin et al. (1992)	Effects of Feedback and Contingent Reinforcement in Reducing Heart Rate Response to Stress	RCT	48	33 min (5 x 3 x 2 min + break 3 min)
15	Larkin et al. (1990)	The Effect of Feedback–Assisted Reduction in Heart Rate Reactivity on Videogame Performance	RCT	67	33 min (5 x 3 x 2 min + break 3 min)
16	Larkin et al. (1989)	Heart Rate Feedback–Assisted Reduction in Cardiovascular Reactivity to a Videogame Challenge	RCT	20	28 min (4 x 6 min + break 4 min)

Table 11.1: List of the selected papers for the review.

11.4 Results

All selected studies presented in Table 11.1 were published in peer–reviewed journals and conferences which have been cited throughout the scientific community. Thus we can state that the research in the BEDS context has been recognized by the scientific community.

11.4.1 Biofeedback

Biofeedback in serious games is a broad subject, but it can be drilled down on a few components: psychophysiology, presentation, and other physiology measures.

Psychophysiology

ECG has been used in most of the investigated papers, measuring HR and HRV (Kim et al., 2013), where the latter has used the frequency domain measures of *the ratio of low and high–frequency powers* (LF/HF) and coherence ratio. Furthermore, respiration and EDA (Kuikkaniemi et al., 2010), also voice analysis (Lara et al., 2018) were used to recognize the emotional state of the players. Such extensive focus on HR and HRV in the BEDS context might have resulted from its strong correlation with arousal, and as a convenient physiological measure of ECG compared to the other physiological signals used to deliver biofeedback. Nevertheless, various signals have been used to deliver biofeedback therapies in serious games. The HR psychophysiology measure has also been correlated with the RSA (Parnandi and Gutierrez-Osuna, 2017) of the deep–breathing skills (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018). Through the BEDS context, evidence shows that biofeedback resulted in higher performance score on the tasks (Zafar et al., 2018; Astor et al., 2013b; Hilborn et al., 2013; Bouchard et al., 2012; Changchun et al., 2009), compared to the control condition playing the game without biofeedback. It has been shown that resonant frequency of breathing influences decision–making performance and ER (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Kim et al., 2013). Furthermore, it has been found there was a significant increase in the mentioned ER physiology indices and a decrease in HR, for the treatment group (Parnandi and Gutierrez-Osuna, 2018; Astor et al., 2013b; Kim et al., 2013; Bouchard et al., 2012; Goodie and Larkin, 2006; Larkin et al., 1989). Moreover, it has been found that using biofeedback resulted in higher attention scores (Kim et al., 2013), while higher HR arousal resulted in worse reaction time (Jerčić et al., 2012).

Presentation

Biofeedback can be presented in many ways but in the context of BEDS they all boil down to two methods: direct visual (overt) and indirect gameplay (covert) presentation. The direct overt presentation presents physiological information directly to the individual (e.g., via a visual display) about their emotional/physiological state using measurable presentation such as a bar (Zafar et al., 2018; Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012); a number (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017); a color of background and game artefacts (Astor et al., 2013b; Jerčić et al., 2012; Larkin et al., 1992); sound (Parnandi and Gutierrez-Osuna, 2018; Lara et al., 2018); and HRV signal (Kim et al., 2013). Moreover, these investigations present the user with the target emotional/physiological state through visual indicators (e.g., message, arrow) (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017), or through occluding the view (Hilborn et al., 2013; Bouchard et al., 2012). The indirect covert presentation presents the physiological information indirectly through subtle changes in gameplay mechanics by changing game difficulty in proportion to the individual's stress levels, for reducing control of the game (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Lara et al., 2018; Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012; Kuikkaniemi et al., 2010; Changchun et al., 2009; Larkin et al., 1992) and through manipulation of the game score (Larkin et al., 1992). The purpose of explicit biofeedback in the field of medicine is to make the individuals more aware of their bodily processes. This is achieved through displaying biofeedback information in a clear and easily perceivable way. Regarding implicit biofeedback, biosignals of the individuals modulate the system and its behavior, such that the individuals may not even become consciously aware of the feedback but still sense it on a subconscious level. Therefore, the individuals have direct and conscious control over the system in the case of explicit biofeedback, while the system is controlled indirectly through the affective signals in the case of implicit biofeedback (Kuikkaniemi et al., 2010). Nevertheless, the separation between the explicit and implicit biofeedback is not perfect since the individuals may realize how the implicit biofeedback modulates the system and thereby gain control over it. In that case, separation becomes blurry, and the implicit biofeedback becomes explicit.

Regarding the BEDS context, direct visual presentation of biofeedback was successful in maintaining the target physiological state during the task (Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017). On another hand, gameplay indirect biofeedback presentation resulted in deep–breathing skill–acquisition and transfer (Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017). The players receiving overt and combined overt–covert biofeedback exhibited greater reductions in Hr and better control to the serious game task, compared to the control group, with the combined ones having a significantly greater effect. This effect was also sustained for the post–training assessment task (Parnandi and Gutierrez-Osuna, 2018; Larkin et al., 1992). These claims might be important waypoints to using these different biofeedback presentations for different purposes, where the target behavior is optimally presented using visual presentation, while the skill–acquisition/transfer is optimally supported using the gameplay presentation of biofeedback. It has been found that each biofeedback presentation supports Er skill, while skill–acquisition of down-regulating high level of arousal leads to a better decision–making performance (Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Astor et al., 2013b; Hilborn et al., 2013; Bouchard et al., 2012; Jerčić et al., 2012), even in the case of brain injured individuals (Kim et al., 2013). Players were found to perform worse in the overt explicit biofeedback presentation condition when they had to regulate arousal through Eda, in contrast to the covert implicit condition. Possibly controlling Eda signal is not as straightforward as the respiration for the biofeedback applications since the decision–making performance increased significantly in the overt explicit respiration condition (Kuikkaniemi et al., 2010).

There is evidence that using the combined biofeedback delivering visual and gameplay biofeedback simultaneously leads to performance increase on the decision–making task in serious games (Kim et al., 2013). Furthermore, such combination of biofeedback presentation might support better skill–acquisition/transfer to other contexts than presenting each one individually (i.e., increased skill–acquisition/transfer of down–regulating high levels of Hr arousal and deep breathing) (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012; Larkin et al., 1992).

Physiology Measures

Measures employed in the selected studies were: Hrv using time measure of *the fraction of consecutive interbeat intervals greater than 50 ms* (pNN50); Eda using skin conductance response which includes short–term phasic, emotional responses to specific stimuli (Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017); and salivary cortisol, Hr, blood pressure and Rsa (Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Bouchard et al., 2012; Larkin et al., 1992). In addition to previously reported findings regarding the Beds context, it has been found that Er achieved stress and arousal reduction using salivary cortisol, Hrv–pNN50, Eda–skin conductance response and breathing rate (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Bouchard et al., 2012). The players receiving overt and combined overt–covert biofeedback exhibited a greater reduction in systolic blood pressure response to the serious game task, compared to the control group, with the combined one having a significantly greater effect (Goodie and Larkin, 2001; Larkin et al., 1990). This effect was also found in the post–training assessment task. On another hand, diastolic blood pressure response revealed no significant effects (Goodie and Larkin, 2001; Larkin et al., 1992).

11.4.2 Tasks

The tasks used in the Beds context can be divided into serious games, decision–making, attention investigation, and arousal validation. Regarding the validation of Er skill–transfer, the 'Halstead Category Test' is well–established and validated with reproducible reliability (Dikmen et al., 1999), and it has been found to have a significant association with problem–solving skills (Reitan and Wolfson, 1996). It has been used as a decision–making task to validate the increase in performance, while the 'Integrated Visual and Auditory Continuous Performance Test' has been used to validate the increase in attention (Kim et al., 2013). Furthermore, performance on the Stroop color–word test and mental arithmetic tasks, King of Math and 'counting backward by N' have been used as well (Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Goodie and Larkin, 2006; Larkin et al., 1992). It has been found that lower Hr responses were observed on the mental arithmetic challenge for the individuals who received the biofeedback treatment in serious games (Goodie and Larkin, 2006).

Performance results on the decision–making task using Er were obtained through a number of measurable outcomes, which can be divided into physiology measurements, questionnaires, and task performance. The performance was measured through biofeedback tasks in the decision–making serious games in all of the selected papers, using task performance score, response time and biofeedback performance score. On another hand, questionnaires measuring performance were: Behavior Rating Inventory of Executive Function (Brief–A–Informant), and The Trainees Evaluation Sheet (Kim et al., 2013; Bouchard et al., 2012). External motivating rewards (i.e., monetary and movie tickets) were used to promote skill–transfer to real–world problems (Astor et al., 2013b; Kim et al., 2013; Hilborn et al., 2013; Jerčić et al., 2012; Changchun et al., 2009; Goodie and Larkin, 2006; Larkin et al., 1992).

Serious game	Arousal	Time–pressure	Difficulty	Fast–paced	Outcomes	Biofeedback presentation	Regulation target	Type
Frozen Bubble	X	X			skill–acquisition	HRV	respiration	mobile
Left4Dead	X		X		decision–making	HR	respiration	PC
Sno–Cat	X				decision–making	HR	HR	PC
Space Investors	X		X		decision–making	HR	HR	PC
Auction Game	X	X	X	X	decision–making	HR	HR	PC
Basic Math		X	X	X	decision–making	VOC		PC
Pong	X	X	X	X	skill–acquisition	EDA	EDA	PC
Emoshooter	X	X	X	X	decision–making	EDA	respiration	PC
Pacman Zen	X	X			skill–acquisition	HRV	respiration	mobile
Dodging Stress	X	X			skill–acquisition	HRV	respiration	mobile

Table 11.2: List of the serious games and their arousal (stress) elicitation methods on a decision–making task, identified in this review. The games listed are: Frozen Bubble (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017), Left4Dead (Bouchard et al., 2012), Sno–Cat (Goodie and Larkin, 2006; Larkin et al., 1992), Space investors (Hilborn et al., 2013), Auction game (Astor et al., 2013b; Jerčić et al., 2012), Basic Math (Bontchev and Vassileva, 2016), Pong (Changchun et al., 2009), Emoshooter (Kuikkaniemi et al., 2010), Pacman Zen (Zafar et al., 2018), Dodging Stress (Zafar et al., 2018).

Serious Games

Serious games aim to present a demanding and stressful task that can elicit sufficient arousal using time pressure, difficulty and fast–paced decisions in all of the selected papers. List of all the identified serious game and their methods of stress and arousal elicitation are shown in Table 11.2. In selected studies, behavior modulation has been controlled using two strategies: NR–IC and contingent reinforcement conditioning.

In the Frozen Bubble, players are presented with an arena containing a spatial arrangement of colored bubbles, and the goal is to clear the arena by placing a bubble next to two or more of the same color which makes them disappear. Through touch screen on an Android device, players control the orientation and firing of a small cannon that shoots bubbles of random color. The ceiling of the arena drops throughout the gameplay, and the game ends after the whole arena fills up. In the Left4Dead players have to collaborate to survive the zombie apocalypse. As part of a team, players have to exit a hideout and reach a rallying point on a farm while hordes of infected zombies are trying to kill them. Navigation is performed with keyboard keys and looking/firing with the mouse. In the Sno–Cat players control a vehicle up a simulated mountain, evading randomly appearing trees. The speed of the vehicle is fixed, while the movement is controlled by a joystick, and the brake is controlled by the fire button. The goal is to maximize the distance traveled, which is influenced by the amount of braking and crashes occurring during the gameplay. In the Space Investor players navigate a spaceship from one planet to another in space, while trying to avoid getting hit by the asteroids. Players cannot die in the game and only lose resources. The spaceship is automatically moving forward through space and is frequently approached by asteroids that must be shot down in order for the ship not to get hit. This is done with the mouse buttons, left for the primary weapon and right for the secondary. In the Auction Game players trade fictional 'goods' on the stock market. The players are presented with a task to calculate a mean value from three given price estimations, to be able to reach a buy or sell decision at the correct price. To make a decision, players have to use a mouse to click on the buy or sell button on the screen. In the Basic Math game, players must choose the correct mathematical operation by moving the falling object, obtaining points by dropping the object into the correct container before it reaches the bottom of the screen. Actions are performed with keyboard keys. In the Pong game players defend their side of the screen with a paddle changing size and speed, against an incoming ball that also changes speed and size. Movements are performed with keyboard keys. In the Emoshooter players have to aim at and shoot as many far–away enemies as possible, which attack the player in big formations. The goal is to keep the character unharmed by evading the bullets and avoiding falling into lava. Navigation is performed with keyboard keys and looking/firing with the mouse. Pacman Zen is similar to the original Pac–Man game where players collect the white dots strewn across the maze. The difference in this version is that every ghost directly chases Pac–Man instead of enacting a more complex chasing behavior. When all the dots have been eaten, players advance to the next level. With each new level, the speed goes up, increasing the difficulty. Furthermore, if Pac–Man comes in contact with a ghost, a life is lost. There are three lives until the game ends. To navigate, players have to use a swiping motion on the mobile screen. In Dodging Stress, players guide the ball from one end of the screen to the other, avoiding obstacles in the way. The goal is to reach the target on the other end of the screen. The player must avoid hitting obstacles that move randomly around the playing field, or the game resets to the starting position. The game also features one continuous level where the game difficulty increases each time players reach the target. To navigate, players have to tilt the mobile device.

11.4.3 Emotion–Regulation

Emotions have been investigated in the Beds context for raising awareness as a prerequisite step for better Er and decision–making performance (Astor et al., 2013b; Jerčić et al., 2012). As somatic hypothesis promotes the embodiment, assessment, and regulation of emotions (Kim et al., 2013), most of the identified papers of the mentioned investigations were focused on how stress during a demanding decision–making task elicits a high level of arousal. It was found that the participants were aware of their arousal during the main biofeedback serious game task, which enabled them to avoid being overpowered by their emotions, and motivated them to develop skills to interpret and down–regulate those affective states of high arousal (Jerčić et al., 2012).

Relaxation

Regarding the Beds context, evidence shows that there was an increase for the treatment group in Er for the downregulation of high arousal (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Astor et al., 2013b; Kim et al., 2013; Bouchard et al., 2012; Larkin et al., 1992). It has been found that using a correct reappraisal Er strategy leads to better performance in the Beds context (Astor et al., 2013b), but there has been no validation that down–regulating high stress and arousal leads to a better performance on the decision–making task, even though this has been a prevalent method in Beds research. Nevertheless, all studies conclude that down–regulating high arousal, compared to the baseline condition, is beneficial for the performance on the decision–making task in serious games, where higher Hrv resulted in better performance and attention scores (Kim et al., 2013).

Skill transfer of Er and down–regulation of high arousal to increase performance has been recognized as an important problem in the field, but it has been investigated only in eight of the identified studies context through tasks (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Goodie and Larkin, 2006; Larkin et al., 1992). Findings from the post–treatment validation tasks show a successful Er skill–transfer where the decision–performance score was increased, while the respiration rate, Hr and Eda responses decreased for the treatment in contrast to the control group (Zafar et al., 2018). These findings gave evidence that the partial reinforcement improves skill–transfer between the tasks (Parnandi and Gutierrez-Osuna, 2018).

11.4.4 Valence

Regarding the Beds context, evidence shows that players were able to experience pleasant valence affect and reduce their unpleasant valence affect (Lara et al., 2018). Results from these investigations indicated that positive affect could improve problem–solving and decision–making abilities (Kim et al., 2018).

Furthermore, it was found that the response time in a decision–making task is slower for players in low valence (unpleasant) affect compared to the other emotional states (Lara et al., 2018).

11.4.5 Healthcare

Certain health conditions have for a consequence reduced decision–making performance connected to the inability to process emotions correctly. The Beds context has emerged as a valuable application of the mentioned concepts on a humanistic problem. It has been found that individuals with the Ans dysregulation (e.g., brain injury, post–traumatic stress disorder) have a limited possibility to modulate their Hr, which in turn results in low amplitudes of the Hrv. This problem has been traced to deficits in executive functioning, which is related to poor decision–making performance on the tasks (Kim et al., 2013; Bouchard et al., 2012). There is an association between successful Er and Hrv biofeedback, where even brain–injured individuals can acquire this skill and build mental resilience, but their executive functioning might not improve through this method (Kim et al., 2013). On another hand, serious games with biofeedback provide a nonpharmacological means to reduce cardiovascular reactions to behavioral tasks, potentially reducing the risk of developing cardiovascular disease or hypertension (Larkin et al., 1989).

11.4.6 Methodology

The approach taken in these studies was to administer Er training in the Beds context and investigate the effects of Er using the same or side validation tasks. The proposed solutions in all of the selected studies were to relate if Hr or Eda can be regulated through different biofeedback presentations (i.e., visual, gameplay and combined) to improve Er skill–acquisition/transfer and increase performance on decision–making tasks (Parnandi and Gutierrez-Osuna, 2017; Bouchard et al., 2012).

All of the selected studies evaluated the effectiveness of biofeedback serious games for Er skill–acquisition through a stressful decision–making task across varied application fields. Moreover, all investigated studies examined objective measures of player arousal, valence and behavior throughout the gameplay. The design adopted in the studies was related to the physiological outcomes related to the performance on the decision–making task. Furthermore, these studies used high–quality experimental designs being used to support the need for evidence from well–controlled studies where outcomes are relatively straightforward to measure. For the Rct studies Anova was used between the treatment and control conditions, while the quasi–experimental design employed a correlation analysis together with repeated–measures Anova. All studies reported recording a pre–treatment baseline, prior to the task training and treatment were administered.

11.5 Discussion

Regarding our key focus on the effects of rewarding and practicing Er through the Beds context, following claims have been found in the investigated works. Participants can regulate their arousal measured with physiology (i.e., Hrv, Hr, salivary cortisol) through combined biofeedback (visual and gameplay) presentation training in the Beds context (Kim et al., 2013; Bouchard et al., 2012). Such individuals who raised awareness of emotions and increased the skill of Er were able to successfully regulate their arousal, which resulted in better performance, reaction time and attention scores on a decision–making task in serious games (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Astor et al., 2013b; Hilborn et al., 2013; Kim et al., 2013; Jerčić et al., 2012; Bouchard et al., 2012). Nevertheless, the valence side of Russell's circumplex model of emotions (Russell, 1980) was underrepresented in the investigated works, correlating with positive as well as negative affective states which might be related to performance in decision–making.

Extensive Er and biofeedback training prior to the experiment with serious games is necessary to resolve the ambiguity where higher decision–making performance scores, compared to the control groups, were reported in some studies (Astor et al., 2013b; Hilborn et al., 2013; Bouchard et al., 2012), while worse ones were reported in the others (Parnandi and Gutierrez-Osuna, 2018; Larkin et al., 1992). The control groups played an easier non–biofeedback version of the game in all of the experiments and showed a constant game score throughout. These versions of the serious games were without the (negative) effects of the biofeedback increasing the task difficulty. Extensive training should become a mandatory requirement to bring all the participants to the same skill level and normalize their performance results. Such training could be supported by other mindfulness approaches known to increase interoception (e.g., different forms of relaxation exercises, yoga, breathing patterns, or mindfulness courses), as a gap in knowledge has been identified of how these other mindfulness approaches affect performance in the Beds context. This would enable researchers to make general claims of using biofeedback in serious games to support Er and increase performance on decision–making tasks. These suggestions could also save discarded datasets because of participants who were not able to follow the instructions on how to perform the experiment. The benefits of practicing Er are largely dose–dependent, which means that the individuals who consistently practice tend to benefit both in the short– and long–term, which is lacking in the current research investigations (Culbert, 2017). Therefore, such environments could be a valuable support for the users and contribute to long–term skill development of effective Er.

Hr is probably a poor measure for a physiological assessment of arousal, as many controversial results regarding it were reported. Some studies reported a decrease in Hr arousal in the Beds context (Astor et al., 2013b), while others reported an increase (Bouchard et al., 2012), and no significant change between groups (Larkin et al., 1992). This controversy is also related to the performance assessment on a decision–making task. Some studies reported that higher Hr arousal results in lower performance (Astor et al., 2013b; Jerčić et al., 2012), while others reported that lower Hr arousal results in lower performance (Parnandi and Gutierrez-Osuna, 2018), and also no significant change between groups (Larkin et al., 1992). Poor selection of a biofeedback presentation through a scoring mechanism was a big factor in the discussion, where the biofeedback treatment group had worse decision–making performance, compared to the non–biofeedback control group. The scoring mechanism has not had a clean separation between the game performance and biofeedback performance score (Goodie and Larkin, 2006; Larkin et al., 1989). Therefore, the players were confused while observing that their Hr biofeedback improved their game score, while at the same time they were not improving on their decision performance. The controversy regarding these studies claims that decision–making serious games in the affective computing field can induce sufficient arousal to participants, such that they can practice Er in such Beds context and be correctly rewarded for it (Astor et al., 2013b). On the contrary, others claim that sufficient arousal was not achieved in such context (Bouchard et al., 2012). The uncertainty to whether higher or lower arousal is correlated with better performance on the decision–making task in serious games might be due to lacking in methodology regarding the validation of serious games. The optimal level of stress and arousal for a given decision–making task depends on the context of the task and the individuals performing on it. Thus it cannot just be said that down–regulation will increase performance. Experiments need to build a player–centered design methodology and validate whether the decision–making tasks in serious games elicited sufficient levels of arousal (Parnandi and Gutierrez-Osuna, 2017; Astor et al., 2013b; Bouchard et al., 2012; Jerčić et al., 2012). Even more so, since arousal is correlated with the performance score through a bell–shaped curve of Yerkes–Dodson–Law, which states that there exists an optimal arousal level for high performance on the task (Parnandi and Gutierrez-Osuna, 2017; Astor et al., 2013b; Bouchard et al., 2012; Jerčić et al., 2012). Without this step, claims cannot be made as to how arousal affects performance on the decision–making task in serious–games. Contrary to Hr, there is evidence that Hrv and salivary cortisol might be more promising measures of arousal.

11.6 Limitations

This review summarizes the research on serious games in the Beds context based on the search terms used within the databases included, without a time period limitation. Therefore, just like other reviews, it does not claim to be comprehensive. Only papers published in English that provided quantitative analysis with a controlled experimental design were considered. Specific focus on Er delivered through biofeedback in the stressful context of decision–making in serious games of this review was used, but the loss of data due to these criteria must be acknowledged. Due to the quantitative focus in the current review, only 16 papers were included, and it is possible that some high–quality qualitative studies were not considered which may inform the development of quantitative measures. Furthermore, research on serious games with biofeedback in other fields is available but was considered outside the scope of this review.

11.7 Future Work

There is a need for involving the view on the Er on a decision–making task from the perspectives of neuroscience, psychophysiology, cognitive psychology which might have a different approach. These areas might give further answers to the cognitive functioning of Er with the biofeedback integration in decision–making serious game tasks, and neural pathway activation between different emotions.

Previous investigations already gave evidence that the Ans is responsible for the executive functioning and decision–making, but the identified investigations did not separate the individual Ans pathways (sympathetic and parasympathetic nervous systems) to tackle the effect of different emotions on performance on a decision–making task. Only arousal has been put into the context of sympathetic and parasympathetic nervous system activation (Astor et al., 2013b; Kim et al., 2013). Moreover, regarding biofeedback measurements, Ecg was used to measure Hr, Hrv, and Rsa (Parnandi and Gutierrez-Osuna, 2018; Astor et al., 2013b; Kim et al., 2013; Hilborn et al., 2013; Jerčić et al., 2012; Bouchard et al., 2012; Larkin et al., 1992). All of these measures are related to the heart, but future should whiteness more diverse investigation of other biofeedback modalities necessary to map out the benefits of biofeedback on Er and decision–making, especially if we want to generalize the claims towards the nervous system, neurology, and psychophysiology.

11.8 Conclusion

The ideas in these selected investigations bring together biofeedback supporting Er skill on decision–making tasks for the novel applications providing evaluation and validation methods for designing serious game artifacts (Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012). The authors tackled this ambiguity in the current research while providing the evidence for the efficiency of different biofeedback presentation modalities (Parnandi and Gutierrez-Osuna, 2017) and how Er skill can be acquired using biofeedback in serious games and transferred to another context through the use of contingent reinforcement (Parnandi and Gutierrez-Osuna, 2018; Bouchard et al., 2012; Larkin et al., 1992). Finally, these investigations gave evidence towards improving the lives of brain–injured individuals (Kim et al., 2013).

The question of whether biofeedback methods increase performance on the decision–making task by supporting Er in serious games, or by reduced arousal making the task easier, is a relevant issue. It was reported that without the proper calibration or normalization of biosignal data the biofeedback effects implemented in serious games would break the game balance (Kuikkaniemi et al., 2010). Unfortunately, these investigations have an ambiguity regarding how arousal influences performance, even if increased performance was reported. More specifically, this ambiguity in the mechanism of biofeedback loop defines that difficulty in the serious games is endogenously connected to arousal, where individuals with higher arousal had a more difficult decision–making task which might have made them more aroused, and vice–versa. Especially since this ambiguity was reflected on the game score, which defines performance on the task. Compared to the control groups, the serious games were probably more difficult also for the individuals using biofeedback on the decision–making tasks, since the extensive Er training was missing and the biofeedback groups had to learn how to regulate Hr, which might have increased arousal on the tasks. The discrepancy in decision–making performance results may be partially attributed to the length of Hr biofeedback training provided in the selected studies especially since the evidence showed that with a sufficient training period, the discrepancy disappeared (Larkin et al., 1990). Control groups should be presented with decision–making tasks without biofeedback, to investigate if the participants used Er skills at all. To make stronger claims, it is also necessary to introduce a sham–biofeedback condition where the game performance score and difficulty are independent of biofeedback, without influencing one another. Moreover, the sham–biofeedback control group should be presented with false biofeedback or biofeedback not connected to the decision–making task, to untangle the biofeedback loop and investigate how emotions and ER affect the decision–making task performance. Otherwise, one group might have a lower performance just because it does not have the necessary information which affects the performance score, which makes it hard to compare results from the control and sham–biofeedback conditions against the treatment groups. Without standardizing and integrating sham–biofeedback and control conditions in the experiment design, as well as screening and validating that the participants practiced ER during the task, general conclusions cannot be drawn on how biofeedback supports ER and affects decision–making performance in general. Most of the investigated studies sport this limitation. Investigations in this area should untangle the biofeedback loop and remove the ambiguity that the task just got easier with reduced arousal, as was attempted in three of the selected investigations (Astor et al., 2013b; Jerčić et al., 2012; Larkin et al., 1990). Control groups have been used in these investigations with: a more difficult gameplay (Parnandi and Gutierrez-Osuna, 2017), less difficult gameplay (Astor et al., 2013b), no biofeedback or task presented (Bouchard et al., 2012) and without biofeedback on the task (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Astor et al., 2013b; Changchun et al., 2009; Goodie and Larkin, 2006; Larkin et al., 1992), compared to a treatment group. Similarly, sham–biofeedback groups have been used in these investigations with: biofeedback presented but not connected to the task (Larkin et al., 1992), and biofeedback connected but not presented on the task (Astor et al., 2013b).

Previous investigations found that the skill–transfer is highly contextual. The BEDS method should be applied to concrete research contexts which include the relevant ecology of the target audience and tasks resembling naturally–occurring stressors. This provides support for the serious games applications that allow framing of a decision–making task in the relevant context. Investigations in the ER have used validation tasks to assess skill–transfer of individuals' abilities to down–regulate high arousal in the absence of biofeedback (Zafar et al., 2018; Goodie and Larkin, 2006; Larkin et al., 1992). Through the investigated works, the ecological validity of participants has been used in the military context (Bouchard et al., 2012) and brain–injury context (Kim et al., 2013), while the rest of the studies have used student participants. Therefore, as the field moves forward, a relevant ecological valid cohort for the participants is necessary to investigate these effects in different contexts. These limitations make it hard to draw general conclusions outside of the presented context.

All of the identified investigations are not–generalizable, context–dependent, without proper disagreeing views being presented. Therefore, exactly these aspects should be the future direction in this area. These investigations provide evidence motivating the different biofeedback presentation modalities in the Beds context (Parnandi and Gutierrez-Osuna, 2017; Larkin et al., 1992). Moreover, they give evidence regarding validation of sufficient arousal elicited which allows the suppression/reappraisal Er strategy to be used to increase performance on the decision–making task (Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012; Bouchard et al., 2012). On another hand, they also raise future questions regarding the ambiguity of higher or lower arousal increasing decision–making performance through biofeedback presentation. Furthermore, they aim towards a more player–centered design of serious games, which motivates their use as a solution for the investigated scientific problems (Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012). Furthermore, these investigations raise future questions regarding skill–transfer to different task contexts (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Hilborn et al., 2013).

Nevertheless, the agreement between the investigated studies claims that the practice of Er using biofeedback improves performance on a decision–making task in serious games through regulating levels of physiological arousal, where combined biofeedback presentation (i.e., direct visual and indirect gameplay) supports such Er skill–acquisition and transfer, even in the case of brain injured individuals (Parnandi and Gutierrez-Osuna, 2018; Zafar et al., 2018; Parnandi and Gutierrez-Osuna, 2017; Zafar et al., 2017; Astor et al., 2013b; Kim et al., 2013; Hilborn et al., 2013; Bouchard et al., 2012; Jerčić et al., 2012; Larkin et al., 1992). Furthermore, designed serious game artifacts have been validated through evaluative and playtesting studies as an effective platform for practicing Er skill–acquisition (Astor et al., 2013b; Hilborn et al., 2013; Jerčić et al., 2012).

Practicing the skill of Er on decision–making in serious games can lead to better decision performance on the task, but careful consideration and standardization regarding methodological approach (i.e., a sham–biofeedback condition, control condition, Er pre–training, biofeedback measures, and presentation modalities) are needed to observe the benefits of such methods.

References

Aarseth, E., Smedstad, S., and Sunnanå, L. (2003). A multidimensional typology of games. In *Level Up: Digital Games Research Conference*, pages 48–53. The European Commission.

Abler, B. and Kessler, H. (2009). Emotion regulation questionnaire–Eine deutschsprachige Fassung des ERQ von Gross und John. *Diagnostica*, 55(3):144–152.

Abt, C. (2002). *Serious games*. University Press Of America.

Acharya, U., Joseph, K., and Kannathal, N. (2006). Heart rate variability: a review. *Medical and Biological Engineering and Computing*, 44(12):1031–1051.

Achtman, R. L., Green, C. S., and Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative neurology and neuroscience*, 26(4-5):435–446.

Adam, M. and Gamer, M. (2011). Measuring emotions in electronic markets. In *Proceedings of the International Conference on Information Systems (ICIS)*, pages 1–19. AIS Electronic Library (AISeL).

Adam, M., Krämer, J., and Weinhardt, C. (2013). Excitement Up! Price Down!; Measuring Emotions in Dutch Auctions. *International Journal of Electronic Commerce (IJEC)*, 17(2):7–40.

Adam, M. T. P., Krämer, J., Jähnig, C., Seifert, S., and Weinhardt, C. (2011). Understanding auction fever: a framework for emotional bidding. *Electronic Markets*, 21(3):197–207.

Adam, M. T. P., Krämer, J., and Müller, M. B. (2015). Auction Fever! How Time Pressure and Social Competition Affect Bidders' Arousal and Bids in Retail Auctions. *Journal of Retailing*, 91(3):468–485.

Agrawal, P., Liu, C., and Sarkar, N. (2008). Interaction between human and robot an affect-inspired approach. *Interaction Studies*, 9(2):230–257.

Akselrod, S., Gordon, D., and Ubel, F. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science*, 213(4504):220–222.

Al Osman, H., Eid, M., and El Saddik, A. (2014). U-biofeedback: A multimedia-based reference model for ubiquitous biofeedback systems. *Multimedia Tools and Applications*, 72(3):3143–3168.

Albert, W. and Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann Pub, 2nd edition.

Aldao, A., Nolen-Hoeksema, S., and Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical psychology review*, 30(2):217–237.

Algoe, S. B. and Fredrickson, B. L. (2011). Emotional Fitness and the Movement of Affective Science From Lab to Field. *American Psychologist*, 66(1):35–42.

Anderson, P. H. and Lawton, L. (2009). Business Simulations and Cognitive Learning: Developments, Desires, and Future Directions. *Simulation and Gaming*, 40(2):193–216.

Andreassi, J. L. (2007). *Psychophysiology: Human behavior and physiological response*. Lawrence Erlbaum Associates, 5th edition.

Andreassi, J. L. (2013). *Psychophysiology: Human behavior and physiological response*. Psychology Press, 5th edition.

Annetta, L. A. (2010). The 'I's' Have It: A Framework for Serious Educational Game Design. *Review of General Psychology*, 14(2):105–112.

Anttonen, J. and Surakka, V. (2005). Emotions and heart rate while sitting on a chair. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 491–499, New York, New York, USA. ACM Press.

Appelhans, B. and Luecken, L. (2006). Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, 10(3):229–240.

Ariely, D. and Loewenstein, G. (2006). The heat of the moment: the effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, 19(2):87–98.

Astor, P., Adam, M., Jähnig, C., and Seifert, S. (2011). Measuring regret: Emotional aspects of auction design. In *Proceedings of the 19th Conference on Information Systems(ECIS 2011)*. AIS Electronic Library (AISeL).

Astor, P. J., Adam, M. T. P., Jähnig, C., and Seifert, S. (2013a). The joy of winning and the frustration of losing: A psychophysiological analysis of emotions in first-price sealed-bid auctions. *Journal of Neuroscience, Psychology, and Economics*, 6(1):14–30.

Astor, P. J., Adam, M. T. P., Jerčić, P., Schaff, K., and Weinhardt, C. (2013b). Integrating Biosignals into Information Systems: A NeuroIs Tool for Improving Emotion Regulation. *Journal of Management Information Systems*, 30(3):247–278.

Bach, D., Friston, K., and Dolan, R. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International Journal of Psychophysiology*, 76(1):52–55.

Bargas-avila, J. a. and Hornbæk, K. (2011). Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2689–2698.

Bartneck, C. and Hoek, M. V. D. (2007). Daisy, Daisy, give me your answer do!: switching off a robot. In *Proceeding of the ACM/IEEE international conference on Human-robot interaction*, pages 217–222, Arlington, VA. ACM Press, New York.

Bashashati, A., Fatourechi, M., Ward, R. K., and Birch, G. E. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of neural engineering*, 4(2):32–57.

Bechara, A. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebral Cortex*, 10(3):295–307.

Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and cognition*, 55(1):30–40.

Bechara, A. and Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, 52(2):336–372.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304):1293–1295.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, 22(3):292–299.

Benedek, M. and Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658.

Bengtsson, B. and Burgoon, J. (1999). The impact of anthropomorphic interfaces on influence understanding, and credibility. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, pages 1–15.

Benikos, N., Johnstone, S. J., and Roodenrys, S. J. (2013). Varying task difficulty in the Go/Nogo task: The effects of inhibitory control, arousal, and perceived effort on ERP components. *International Journal of Psychophysiology*, 87(3):262–272.

Bergman, J. S. and Johnson, H. J. (1972). Sources of Information Which Affect Training and Raising of Heart Rate. *Psychophysiology*, 9(1):30–39.

Berntson, G. and Cacioppo, J. (1994). Autonomic cardiac control. III. Psychological stress and cardiac response in autonomic space as revealed by pharmacological blockades. *Psychophysiology*, 31(6):599–608.

Berntson, G. G., Quigley, K. S., and Lozano, D. (2007). Cardiovascular psychophysiology. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology*, chapter 3, pages 182–210. Cambridge Univ. Press, Cambridge.

Bethel, C. L., Salomon, K., Murphy, R. R., and Burke, J. L. (2007). Survey of psychophysiology measurements applied to human-robot interaction. In *Robot and Human Interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 732–737. IEEE.

Biais, B., Hilton, D., Mazurier, K., and Pouget, S. (2005). Judgemental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market. *Review of Economic Studies*, 72(2):287–312.

Björk, S. and Holopainen, J. (2004). *Patterns in game design*. Charles River Media, Hingham.

Bontchev, B. and Vassileva, D. (2016). Assessing engagement in an emotionally-adaptive applied game. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM '16*, pages 747–754. ACM.

Bos, D. P.-o., Reuderink, B., Laar, B. V. D., Gürkök, H., Mühl, C., Poel, M., Nijholt, A., and Heylen, D. (2010). Brain-Computer Interfacing and Games. In *Brain-Computer Interfaces*, pages 149–178. Springer.

Bouchard, S., Bernier, F., Boivin, É., Morin, B., and Robillard, G. (2012). Using biofeedback while immersed in a stressful videogame increases the effectiveness of stress management skills in soldiers. *PLoS ONE*, 7(4).

Boucsein, W. (2012). *Electrodermal activity*. Springer, 2 edition.

Boutcher, Y. N. and Boutcher, S. H. (2006). Cardiovascular response to Stroop: effect of verbal response and task difficulty. *Biological psychology*, 73(3):235–241.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., and Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers and Education*, 94:178–192.

Bradley, M. and Lang, P. (2000). Emotion and motivation. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology*, chapter 22, pages 602–642. Cambridge Univ. Press, 2 edition.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, 1(3):276–298.

Bradley, M. M., Miccoli, L., Escrig, M. a., and Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607.

Braithwaite, J. and Watson, D. (2013). A Guide for Analysing Electrodermal Activity (EDA) and Skin Conductance Responses (SCRs) for Psychological Experiments. *Psychophysiology*, 49(1):1017–1034.

Breazeal, C. (2002). *Designing sociable robots*. The MIT Press.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Brooks, A., Gray, J., and Hoffman, G. (2004). Robot's play: interactive games with sociable machines. *Computers in Entertainment*, 2(3):1–18.

BSA and FSA (2002). Adult Financial Capability Framework. Technical report, Financial Services Authority and Basic Skills Agency, London.

Buckley, K. E. and Anderson, C. A. (2006). A Theoretical Model of the Effects and Consequences of Playing Video Games. In Vorderer, P., Bryant, J., and Zillmann, D., editors, *Playing Video Games - Motives, Responses, and Consequences*, chapter 24, pages 363–378. LEA, Mahwah, NJ.

Burgoon, J. and Bonito, J. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior*, 6(16):553–574.

Burgoon, J. and Bonito, J. (2002). Testing the interactivity principle: Effects of mediation, propinquity, and verbal and nonverbal modalities in interpersonal interaction. *Journal of Communication*, 52(3):657–677.

Butler, J. and Agah, A. (2001). Psychological effects of behavior patterns of a mobile personal robot. *Autonomous Robots*, 10(2):185–202.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., and Ito, T. A. (2000). The psychophysiology of emotion. In *Handbook of Emotions*, chapter 11, pages 173–191. Guilford Press, 2nd edition.

Cacioppo, J. T., Petty, R. E., Losch, M. E., and Kim, H. S. (1986). Electromyographic Activity Over Facial Muscle Regions Can Differentiate the Valence and Intensity of Affective Reactions. *Journal of Personality and Social Psychology*, 50(2):260–268.

Cacioppo, J. T. and Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *The American psychologist*, 45(1):16–28.

Caillois, R. (2001). *Man, play, and games.* University of Illinois Press.

Calvo, R. R. and D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and their Applications. *Affective Computing, IEEE Transactions*, 1(1):18–37.

Cannon, B. (1994). Walter Bradford Cannon: Reflections on the Man and His Contributions. *International Journal of Stress Management*, 1(2):145–158.

Cannon-Bowers, J. A. and Salas, E. E. (1998). *Making decisions under stress: Implications for individual and team training.* American psychological association.

Caria, A., Veit, R., Sitaram, R., Lotze, M., Weiskopf, N., Grodd, W., and Birbaumer, N. (2007). Regulation of anterior insular cortex activity using real-time fMRI. *NeuroImage*, 35(3):1238–1246.

Carney, R. and Saunders, R. (1995). Association of depression witk reduced heart rate variability in coronary artery disease. *The American Journal of Cardiology*, 76(8):562–564.

Carter, M., Downs, J., and Nansen, B. (2014). Paradigms of games research in HCI: a review of 10 years of research at CHI. In *Proceedings of the 2014 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY 2014)*, pages 27–36. ACM.

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and McOwan, P. W. (2012). Detecting engagement in hri: An exploration of social and task-based context. In *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, pages 421–428. IEEE.

Chanel, G. (2009). *Emotion assessment for affective computing based on brain and peripheral signals*. PhD thesis, University of Geneva.

Changchun, L., Agrawal, P., Sarkar, N., Shuo, C., Liu, C., and Chen, S. (2009). Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback. *International Journal of Human-Computer Interaction*, 25(6):506–529.

Charles, D. and Black, M. (2004). Dynamic Player Modelling: A Framework for Player-centred Digital Games. In *Proceedings of 5th International Conference on Computer Games: Artificial Intelligence, Design and Education (CGAIDE'04)*, pages 29–35.

Charles, D., Kerr, A., and McNeill, M. (2005). Player-centred game design: Player modelling and adaptive digital games. *The Digital Games*, 285(6):285–298.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, 50(4):31–34.

Chiou, A. and Wong, K. W. (2008). Player adaptive entertainment computing (PAEC): Mechanism to model user satisfaction by using neuro linguistic programming (NLP) techniques. In *2008 IEEE Symposium on Computational Intelligence and Games, CIG 2008*, pages 343–349.

Choi, M. and Steptoe, A. (1982). Instructed heart rate control in the presence and absence of a distracting task: The effects of biofeedback training. *Biofeedback and Self-Regulation*, 7(3):257–268.

Clemens, W. J. and Shattock, R. J. (1979). Voluntary Heart Rate Control During Static Muscular Effort. *Psychophysiology*, 16(4):327–332.

Clifford, G. (2007). ECG Statistics, Noise, Artifacts, and Missing Data. In *Advanced Methods and Tools for ECG Data Analysis*, chapter 3, pages 55–100. Artech House, Boston.

Clough, G., Conole, G., and Scanlon, E. (2009). Behavioural finance and immersive games: a pan-European framework for design and evaluation. In *Same places, different spaces. Proceedings ascilite.*

Clough, G., Conole, G., and Scanlon, E. (2010). Using participatory evaluation to support collaboration in an interdisciplinary context. In *Proceedings of the Seventh International Conference on Networked Learning 2010.*

Cohen, J. (1988). Statistical power analysis for the behavioral sciences.

Cohen, R. A. (2011). Yerkes–Dodson Law. In *Encyclopedia of clinical neuropsychology*, chapter 12, pages 2737–2738. Springer.

Cole, P. M., Martin, S. E., and Dennis, T. A. (2004). Emotion Regulation as a Scientific Construct : Methodological Challenges and Directions for Child Development Research. *Child Development*, 75(2):317–333.

Conati, C. and Chabbal, R. (2003). A study on using biometric sensors for monitoring user emotions in educational games. In *Workshop on Assessing and Adapting to User Attitudes and Affect: Why, When and How.*

Connolly, T. M., Boyle, E. A., Macarthur, E., Hainey, T., and Boyle, J. M. (2012). Computers & Education A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2):661–686.

Cooney, M., Nishio, S., and Ishiguro, H. (2014). Affectionate Interaction with a Small Humanoid Robot Capable of Recognizing Social Touch Behavior. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 4(4):1–32.

Cornett, S. (2004). The usability of massively multiplayer online roleplaying games. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, pages 703–710. ACM.

Corti, K. (2006). Games-based Learning; a serious business application. *Informe de PixelLearning*, 34(6):1–20.

Creswell, J. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE Publications, 3rd edition.

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature neuroscience*, 7(2):189–195.

Crone, E. a., Somsen, R. J. M., Van Beek, B., and Van Der Molen, M. W. (2004). Heart rate and skin conductance analysis of antecendents and consequences of decision making. *Psychophysiology*, 41(4):531–540.

Crown Copyright (2003). *An introduction to health and safety.* HSE Books.

Csikszentmihalyi, M. (2008). *Flow: The Psychology of Optimal Experience.* Harper Perennial Modern Classics, New York, US, 1st edition.

Culbert, T. (2017). Perspectives on Technology-Assisted Relaxation Approaches to Support Mind-Body Skills Practice in Children and Teens: Clinical Experience and Commentary. *Children*, 4(20):1–15.

Dautenhahn, K. (2007). Methodology and themes of human-robot interaction: a growing research field. *International Journal of Advanced Robotic Systems*, 4(1):103–108.

Davidson, R. J. (2003). Affective neuroscience and psychophysiology: toward a synthesis. *Psychophysiology*, 40(5):655–665.

Davis, M. J. and Addis, M. E. (1999). Predictors of attrition from behavioral medicine treatments. *Annals of Behavioral Medicine*, 21(4):339–349.

Dawson, M., Schell, A., and Filion, D. (2007). The Electrodermal System. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology*, chapter 8, pages 200–223. Cambridge Univ. Press, 2 edition.

Dawson, M. E., Schell, A. M., and Courtney, C. G. (2011). The skin conductance response system, anticipation, and decision-making. *Journal of Neuroscience, Psychology, and Economics*, 4(2):111–116.

De Sonneville, L. M. J., Verschoor, C., Njiokiktjien, C., Op het Veld, V., Toorenaar, N., and Vranken, M. (2002). Facial identity and facial emotions: speed, accuracy, and processing strategies in children and adults. *Journal of clinical and experimental neuropsychology*, 24(2):200–213.

Dehais, F., Sisbot, E. A., Alami, R., and Causse, M. (2011). Physiological and subjective evaluation of a human-robot object hand-over task. *Applied Ergonomics*, 42(6):785–791.

Delaney, J. and Brodie, D. (2000). Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. *Perceptual and motor skills*, 91(2):515–524.

Delplanque, S. and Grandjean, D. (2008). Emotional processing of odors: evidence for a nonlinear relation between pleasantness and familiarity evaluations. *Chemical Senses*, 33(5):469–479.

Demaree, H. and Everhart, D. (2004). Healthy high-hostiles: Reduced parasympathetic activity and decreased sympathovagal flexibility during negative emotional processing. *Personality and Individual Differences*, 36(2):457–469.

Di Stasi, L. L., Antolí, A., Gea, M., and Cañas, J. J. (2011). A neuroergonomic approach to evaluating mental workload in hypermedia interactions. *International Journal of Industrial Ergonomics*, 41(3):298–304.

Dick, P. K. (1968). *Do Androids Dream of Electric Sheep.* Ballantine Books.

Dikmen, S. S., Heaton, R. K., Grant, I., and Temkin, N. R. (1999). Test-retest reliability and practice effects of expanded Halstead- Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, 5(4):346–356.

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology*, 27(5):481–494.

Dimoka, A. and Banker, R. (2012). On the use of neurophysiological tools in IS research: developing a research agenda for NeuroIS. *MIS Quarterly*, 36(3):679–702.

Dimoka, A., Pavlou, P., and Davis, F. (2011). Research Commentary—NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. *Information Systems Research*, 22(4):687–702.

Djajadiningrat, T., Geurts, L., Munniksma, P. R., Christiaansen, G., and de. Bont, J. (2009). Rationalizer: an emotion mirror for online traders. *Proceedings of the International Workshop on Design and Semantics of Form and Movement 5*, pages 39–48.

D'Mello, S., Dowell, N., and Graesser, A. (2013). Unimodal and Multimodal Human Perception of Naturalistic Non-Basic Affective States during Human-Computer Interactions. *IEEE Transactions on Affective Computing*, 4(4):452–465.

Domjan, M. (2014). *The principles of learning and behavior*. Nelson Education.

Doshi, A., Patrick, K., Sallis, J. F., and Calfas, K. (2003). Evaluation of physical activity web sites for use of behavior change theories. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*, 25(2):105–111.

Drachen, A., El-nasr, M. S., and Canossa, A. (2013). Game Analytics - The Basics. In *Game Analytics*, chapter 2, pages 13–40. Springer.

Drachen, A., Nacke, L. E., Yannakakis, G., and Pedersen, A. L. (2010). Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games - Sandbox '10*, pages 49–54.

Driskell, J. E. and Johnston, J. H. (1998). Stress exposure training. In Cannon-Bowers, J. and Salas, E., editors, *Making decisions under stress: Implications for individual and team training*, chapter 8, pages 191–217. American Psychological Association, Washington, DC, USA.

Dror, I. (2008). Technology enhanced learning: The good, the bad, and the ugly. *Pragmatics and Cognition*, 16(2):215–223.

Dunwell, I., Freitas, S. D., and Jarvis, S. (2010). Four-dimensional consideration of feedback in serious games. In *Digital games and learning*, chapter 4, pages 42–62. Continuum Publishing.

Efron, B. and Hastie, T. (1997). Least angle regression. *The Annals of statistics*, 32(2):407–499.

Ekman, P. (1999). Basic Emotions. In Dalgleish, T. and Power, T., editors, *Handbook of cognition and emotion*, chapter 3, pages 45–60. John Wiley and Sons, Ltd, Sussex, UK.

Ekman, P., Levenson, R., and Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210.

El-Nasr, M. S., Drachen, A., and Canossa, A. (2013). *Game analytics: Maximizing the value of player data.* Springer, London, 2013 edition.

Ellis, C. J. (1981). The pupillary light reflex in normal subjects. *The British journal of ophthalmology*, 65(11):754–759.

Ellis, H., Heppell, S., Kirriemuir, J., Krotoski, A., and Mcfarlane, A. (2006). Unlimited Learning: Computer and Video Games in the Learning Landscape. Technical report, Entertainment and Leisure Software Publishers Association and Press enquiries, London.

Esau, N., Kleinjohann, L., and Kleinjohann, B. (2008). Integrating emotional competence into man-machine collaboration. In *IFIP International Conference on Biologically Inspired Collaborative Computing*, volume 268, pages 187–198. Springer.

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2):133–145.

Feigenbaum, E. A. and Simon, H. A. (1962). Serial position. *British Journal of Psychology*, 53(3):307–320.

Fenton-O'Creevy, M., Conole, G., Lins, J. T., Peffer, G., Adam, M. T. P., Lindley, C. A., Clough, G., and Scanlon, E. (2012a). A learning design to support the emotion regulation of investors. In *OECD-SEBI International Conference on Investor Education*, pages 1–16. The Open University.

Fenton-O'Creevy, M., Lins, J. T., Vohra, S., Richards, D. W., Davies, G., and Schaaff, K. (2012b). Emotion regulation and trader expertise: Heart rate variability on the trading floor. *Journal of Neuroscience, Psychology, and Economics*, 5(4):227–237.

Fenton-O'Creevy, M., Nicholson, N., Soane, E., and Willman, P. (2003). Trading on illusions: Unrealistic perceptions of control and trading performance. *Journal of Occupational and Organizational Psychology*, 76(1):53–68.

Fenton-O'Creevy, M., Soane, E., Nicholson, N., and Willman, P. (2011). Thinking, feeling and deciding: the influence of emotions on the decision making and performance of traders. *Journal of Organizational Behavior*, 32(8):1044–1061.

Fiore, S. M., Wiltshire, T. J., Lobato, E. J. C., Jentsch, F. G., Huang, W. H., and Axelrod, B. (2013). Toward understanding social cues and signals in human-robot interaction: Effects of robot gaze and proxemic behavior. *Frontiers in Psychology*, 4:1–15.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143– 166.

Fredrickson, B. L. (2001). The Role of Positive Emotions in Positive Psychology. *American Psychologist*, 56(3):218–226.

Fredrickson, B. L., Mancuso, R. a., Branigan, C., and Tugade, M. M. (2000). The Undoing Effect of Positive Emotions. *Motivation and Emotion*, 24(4):237–258.

Friedenberg, J. and Silverman, G. (2005). *Cognitive science: An introduction to the study of mind*. Sage Publications, Inc, 1 edition.

Friedman, B., Thayer, J., and Tyrrell, R. (1996). Spectral characteristics of heart period variability during cold face stress and shock avoidance in normal subjects. *Clinical Autonomic Research*, 6(3):147–152.

Froment, S., Ginibre, M., Mader, S., Sarafian, A., Schwartz, A., Soriano, D., Topol, A., and Dupire, J. (2009). AZ66: How Can We Play with Emotions? In *International Conference on Entertainment Computin (ICEC)*, volume 5709, pages 327–328. Springer.

Furlan, R., Guzzetti, S., Crivellaro, W., and Dassi, S. (1990). Continuous 24-hour assessment of the neural regulation of systemic arterial pressure and RR variabilities in ambulant subjects. *Circulation*, 81(2):537–547.

Gallagher, S. (2007). Social cognition and social robots. *Pragmatics & Cognition*, 15(3):435–453.

Galluzzi, S., Nicosia, F., Geroldi, C., Alicandri, A., Bonetti, M., Romanelli, G., Zulli, R., and Frisoni, G. B. (2009). Cardiac autonomic dysfunction is associated with white matter lesions in patients with mild cognitive impairment.

Journals of Gerontology - Series A Biological Sciences and Medical Sciences, 64(12):1312–1315.

Geisler, F., Vennewald, N., Kubiak, T., and Weber, H. (2010). The impact of heart rate variability on subjective well-being is mediated by emotion regulation. *Personality and Individual Differences*, 49(7):723–728.

Gendolla, G. and Krüsken, J. (2002). Mood state, task demand, and effort-related cardiovascular response. *Cognition and Emotion*, 16(5):577–603.

Gigerenzer, G. (2002). *Bounded rationality: The adaptive toolbox.* The MIT Press.

Giles, D., Draper, N., and Neil, W. (2016). Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. *European Journal of Applied Physiology*, 116(3):563–571.

Gilleade, K., Dix, A., and Allanson, J. (2005). Affective Videogames and Modes of Affective Gaming: Assist Me, Challenge Me, Emote Me. In *Proceeding of DIGRA 2005*, pages 1–7.

Glanz, K., Rimer, B., and Viswanath, K. (2008). *Health behavior and health education: theory, research, and practice.* John Wiley and Sons.

Goetz, J. and Kiesler, S. (2002). Cooperation with a robotic assistant. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*, pages 578– 579. ACM.

Goetz, J., Kiesler, S., and Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Proceedings of the 12th IEEEWorkshop on Robot and Human Interactive Communication (RO-MAN 2003)*, pages 55–60, San Francisco, CA.

Goldstein, A. (1996). Out of Control : Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, 65(3):272–292.

Goodie, J. L. and Larkin, K. T. (2001). Changes in hemodynamic response to mental stress with heart rate feedback training. *Applied Psychophysiology Biofeedback*, 26(4):293–309.

Goodie, J. L. and Larkin, K. T. (2006). Transfer of heart rate feedback training to reduce heart rate response to laboratory tasks. *Applied Psychophysiology Biofeedback*, 31(3):227–242.

Goodrich, M. a. and Schultz, A. C. (2007). Human-Robot Interaction: A Survey. *Foundations and trends in human-computer interaction*, 1(3):203–275.

Grandey, A. (2000). Emotion regulation in the workplace: A new way to conceptualize emotional labor. *Journal of occupational health psychology*, 5(1):95–110.

Granholm, E. and Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, 52(1):1–6.

Greiner, B. (2004). An online recruitment system for economic experiments. *Forschung und wissenschaftliches Rechnen*, pages 79–93.

Gross, J. (2009). *Handbook of emotion regulation*. The Guilford Press.

Gross, J. J. (1998a). Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression, and physiology. *Journal of personality and social psychology*, 74(1):224–237.

Gross, J. J. (1998b). The Emerging Field of Emotion Regulation : An Integrative Review. *Review of General Psychology*, 2(3):271–299.

Gross, J. J. (2008). Emotion regulation. In Lewis, M., Haviland-Jones, J. M., and Barrett, L. F., editors, *Handbook of emotions*, chapter 31, pages 497–512. The Guilford Press, New York, N.Y., 3rd edition.

Gross, J. J. and John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2):348–362.

Grossman, D. and Christensen, L. W. (2007). *On combat: The psychology and physiology of deadly conflict in war and in peace*. PPCT Research Publications Belleville, IL.

Guardini, P. and Maninetti, P. (2013). Better Game Experience Through Game Metrics: A Rally Videogame Case Study. In *Game Analytics*, pages 325–361. Springer.

Haag, A., Goronzy, S., Schaich, P., and Williams, J. (2004). Emotion recognition using bio-sensors: First steps towards an automatic system. In André, E., Dybkjæ r, L., Minker, W., and Heisterkamp, P., editors, *Tutorial and research workshop on affective dialogue systems*, pages 36–48. Springer Berlin Heidelberg.

Hagelbäck, J., Hilborn, O., Jerčić, P., Johansson, S., Lindley, C., Svensson, J., and Wen, W. (2014). Psychophysiological Interaction and Empathic Cognition for Human-Robot Cooperative Work (PsyIntEC). In *Gearing up and accelerating cross-fertilization between academic and industrial robotics research in Europe*, chapter 5, pages 283–299. Springer International Publishing.

Hamilton, P. (2002). Open source ECG analysis software documentation. *Computers in Cardiology*, 29:101–104.

Hansen, A. L., Johnsen, B. H., and Thayer, J. F. (2003). Vagal influence on working memory and attention. *International Journal of Psychophysiology*, 48(3):263–274.

Hartman, J. and Galati, M. (2000). A revised business game for use in teaching engineering economy or operations management. In *Proceedings of ASEE annual conference proceedings: engineering education beyond the millennium*, pages 18–21. ASEE.

Hebb, D. (1955). Drives and the CNS (conceptual nervous system). *Psychological review*, 62(4):243–254.

Heerink, M., Ben, K., Evers, V., and Wielinga, B. (2008). The influence of social presence on acceptance of a companion robot by older people. *Journal of Physical Agents*, 2(2):33–40.

Heilman, R. M., Crişan, L. G., Houser, D., Miclea, M., and Miu, A. C. (2010). Emotion regulation and decision making under risk and uncertainty. *Emotion*, 10(2):257–265.

Hess, E. H. and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143 (3611)(1964):1190–1192.

Hilborn, O., Cederholm, H., Eriksson, J., and Lindley, C. (2013). A biofeedback game for training arousal regulation during a stressful task: The space investor. In *Proceedings of 15th International conference, HCI International 2013*, pages 403–410. Springer Berlin / Heidelberg.

Hinds, P., Roberts, T., and Jones, H. (2004). Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1):151–181.

Hirsch, J. and Bishop, B. (1981). Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. *American Journal of Physiology-Heart and Circulatory Physiology*, 241(4):620–629.

Hirth, J., Schmitz, N., and Berns, K. (2011). Towards social robots: Designing an emotion-based architecture. *International Journal of Social Robotics*, 3(3):273–290.

Hjelm, S. (2003). Research+ design: the making of Brainball. *Interactions*, 10(1):26–34.

Hocine, N. and Gouaich, A. (2013). Difficulty and Scenario Adaptation: An Approach to Customize Therapeutic Games. In *Serious games for healthcare: Applications and implications*, chapter 6, pages 107–134. IGI Global, Hershey, PA, 1st edition.

Höök, K., Bullock, A., Paiva, A., and Vala, M. (2003). Fantasya and sentoy. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, pages 804–805, New York, New York, USA. ACM Press.

Houghton, A. and Gray, D. (2014). *Making sense of the ECG: a hands-on guide*. CRC Press, 4th edition.

Hu, Y., Wang, D., Pang, K., Xu, G., and Guo, J. (2015). The effect of emotion and time pressure on risk decision-making. *Journal of Risk Research*, 18(5):637–650.

Hudlicka, E. (2008). Affective Computing for Game Design. In *Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation (GAMEON-NA)*, pages 5–12, Montreal, Canada. McGill.

Hugdahl, K. (1995). *Psychophysiology: The mind-body perspective*. Harvard University Press.

Hughes, J. and Stoney, C. (2000). Depressed mood is related to high-frequency heart rate variability during stressors. *Psychosomatic Medicine*, 62(6):796–803.

Hung, J. C. S., Chiang, K. H., Huang, Y. H., and Lin, K. C. (2017). Augmenting teacher-student interaction in digital learning through affective computing. *Multimedia Tools and Applications*, 76(18):18361–18386.

Iancovici, T. V., Osorio, S., and Rosario, B. J. (2011). Biofeedback in Virtual Reality Applications and Gaming. Technical report, University of Massachusetts.

Ijsselsteijn, W., Harper, B., and Group, T. (2000). Virtually there? A vision on presence research. Technical report, Eindhoven University of Technology.

Isbister, K. and Schaffer, N. (2008). *Game usability: advancing the player experience*. CRC Press.

Jaakkola, T. S. and Jordan, M. I. (1997). A variational approach to Bayesian logistic regression models and their extensions. In *Sixth International Workshop on Artificial Intelligence and Statistics*, volume 82, pages 1–12. Springer.

Jain, D., Joska, T., Lee, F., and Burg, M. (2001). Day-to-day reproducibility of mental stress-induced abnormal left ventricular function response in patients with coronary artery disease and its relationship to. *Journal of Nuclear Cardiology*, 8(3):347–355.

James, W. (1884). What Is an Emotion? *Mind*, 19:188–205.

Järvinen, A. (2009). Understanding video games as emotional experiences. In *The video game theory reader*, chapter 2, pages 85–108. Taylor and Francis.

Jerath, R., Edry, J. W., Barnes, V. A., and Jerath, V. (2006). Physiology of long pranayamic breathing: Neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Medical Hypotheses*, 67(3):566–571.

Jercic, P. (2010). The Future of Brain-Computer Interface for Games and Interaction Design. In *Biosplay workshop at Fun and Games Conference*.

Jerčić, P., Sennersten, C., and Lindley, C. (2018a). Modeling cognitive load and physiological arousal through pupil diameter and heart rate. *Multimedia Tools and Applications*.

Jerčić, P., Wen, W., Hagelbäck, J., and Sundstedt, V. (2018b). The Effect of Emotions and Social Behavior on Performance in a Collaborative Serious Game Between Humans and Autonomous Robots. *International Journal of Social Robotics*, 10(1):115–129.

Jerčić, P., Astor, P. J., Adam, M., Hilborn, O., Schaff, K., Lindley, C. A., Sennersten, C., and Eriksson, J. (2012). A Serious Game using Physiological Interfaces for Emotion Regulation Training in the context of Financial Decision-Making. In *Proceedings of the 20th European Conference on Information Systems (ECIS 2012)*. AIS Electronic Library (AISeL).

Johnston, D. W. and Lethem, J. (1981). The Production of Specific Decreases in Interbeat Interval and the Molor Skills Analogy. *Psychophysiology*, 18(3):288–300.

Jonassen, D., Howland, J., Moore, J., and Marra, R. (2003). Learning to solve problems with technology: A constructivist perspective. Technical report, Columbus.

Jones, D. A. and Chang, M. (2011). Educational Massively Multiplayer Online Role Playing Game for Teaching Youth Finance. In *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*, pages 221–223.

Jung, Y. and Lee, K. (2004). Effects of physical embodiment on social presence of social robots. In *Proceedings of The Seventh International Workshop on Presence*, pages 80–87, Spain.

Kabat-zinn, J., Massion, A., Kristeller, J., Peterson, L. G., Fletcher, E., Pbert, L., Lenderking, R., and Santorelli, S. F. (1992). Effectiveness of a Meditation-

Based Stress Reduction Program in the Treatment of Anxiety Disorders. *American Journal of Psychiatry*, 149(7):936–943.

Kahneman, D. and Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 145(3756):1583–1585.

Kahneman, D. and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases, The Psychology of Intuitive Judgement*, 49:1–30.

Kamath, M. and Fallen, E. (1995). Correction of the heart rate variability signal for ectopics and missing beats. In Malik, M. and Camm, A., editors, *Heart rate variability*, chapter 4, pages 75–85. Armonk: Futura.

Kampenes, V. B., Dybå, T., Hannay, J. E., and K. Sjøberg, D. I. (2009). A systematic review of quasi-experiments in software engineering. *Information and Software Technology*, 51(1):71–82.

Kanda, T. (2007). Field trial approach for communication robots. In *Proc. 16th IEEE International Symposium on Robot and Human interactive Communication (RO- MAN 2007)*, pages 665–666.

Kannan, P. and Kopalle, P. (2001). Dynamic pricing on the internet: importance and implications for consumer behavior. *International Journal of Electronic Commerce*, 5(3):63–83.

Karavidas, M. K., Lehrer, P. M., Vaschillo, E., Vaschillo, B., Marin, H., Buyske, S., Malinovsky, I., Radvanski, D., and Hassett, A. (2007). Preliminary results of an open label study of heart rate variability biofeedback for the treatment of major depression. *Applied Psychophysiology Biofeedback*, 32(1):19–30.

Karwowski, W. and Rahimi, M. (1991). Worker selection of safe speed and idle condition in simulated monitoring of two industrial robots. *Ergonomics*, 34(5):531–546.

Kato, P. M. (2010). Video games in health care: Closing the gap. *Review of General Psychology*, 14(2):113–121.

Kelly, H., Howell, K., Glinert, E., Holdig, L., Swain, C., Burrowbridge, A., and Roper, M. (2007). How to build serious games. *Communications of the ACM*, 50(7):44–49.

Kemp, A., Quintana, D., and Gray, M. (2011). Is heart rate variability reduced in depression without cardiovascular disease? *Biological Psychiatry*, 69(4):e3–e4.

Kempson, E., Collard, S., and Moore, N. (2006). Measuring financial capability: An exploratory study for the Financial Services Authority. *Consumer Financial Capability: Empowering European Consumers*, 39.

Kettunen, J., Ravaja, N., and Näätänen, P. (1998). The synchronization of electrodermal activity and heart rate and its relationship to energetic arousal: a time series approach. *Biological Psychology*, 48(3):209–225.

Kidd, C. (2003). *Sociable robots: The role of presence and task in human-robot interaction*. PhD thesis, Massachusetts Institute of Technology.

Kidd, C. and Breazeal, C. (2004). Effect of a robot on user perceptions. In *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robotics and Systems*, volume 4, pages 3559–3564, Japan. IEEE.

Kiesler, S. and Hinds, P. (2004). Introduction to this special issue on human-robot interaction. *Human-Computer Interaction*, 19(1):1–8.

Kim, H.-R. and Kwon, D.-S. (2010). Computational Model of Emotion Generation for Human–Robot Interaction Based on the Cognitive Appraisal Theory. *Journal of Intelligent and Robotic Systems*, 60(2):263–283.

Kim, S., Rath, J. F., Zemon, V., Cavallo, M. M., McCraty, R., Sostre, A., and Foley, F. W. (2018). Problem solving, biofeedback, and severe brain injury: The moderating role of positive affect. *Rehabilitation Psychology*, 63(1):148–154.

Kim, S., Zemon, V., Cavallo, M. M., Rath, J. F., McCraty, R., and Foley, F. W. (2013). Heart rate variability biofeedback, executive functioning and chronic brain injury. *Brain Injury*, 27(2):209–222.

Kivikangas, M. J., Nacke, L., and Ravaja, N. (2011). Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character. *Entertainment Computing*, 2(1):11–16.

Kluger, A. and DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, 119(2):254–284.

Koay, K. (2005). Methodological issues using a comfort level device in human-robot interactions. In *Proc. IEEE International Workshop on Robot and Human Interactive Communication (ROMAN 2005)*, pages 359–364. IEEE.

Koda, T. and Maes, P. (1996). Agents with faces: The effect of personification. In *Fifth IEEE International Workshop on Robot and Human Communication*, pages 189–194. IEEE.

Koelsch, S. and Remppis, A. (2007). A cardiac signature of emotionality. *European Journal of Neuroscience*, 26(11):3328–3338.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., Brooks, D. J., Bench, C. J., and Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, 393(6682):266–268.

Ku, G., Malhotra, D., and Murnighan, J. K. (2005). Towards a competitive arousal model of decision-making: A study of auction fever in live and Internet auctions. *Organizational Behavior and Human Decision Processes*, 96(2):89–103.

Kuhnen, C. M. and Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, 47(5):763–770.

Kuikkaniemi, K., Laitinen, T., Turpeinen, M., Saari, T., Kosunen, I., and Ravaja, N. (2010). The influence of implicit and explicit biofeedback in first-person shooter games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, volume 2, pages 859–868, Atlanta, GA, United states. ACM.

Kulic, D. and Croft, E. (2006). Estimating robot induced affective state using hidden markov models. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 257–262. IEEE.

Laeng, B., Sirois, S., and Gredeback, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, 7(1):18–27.

Lane, R., McRae, K., Reiman, E., and Chen, K. (2009). Neural correlates of heart rate variability during emotion. *Neuroimage*, 44(1):213–222.

Lang, P., Bradley, M., and Cuthbert, B. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* National Institute of Mental Health C enter for the Study of Emotion and Attention, Bethesda, MD.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American psychologist*, 50(5):372–385.

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273.

Lang, P. J. and Twentyman, C. T. (1976). Learning to Control Heart Rate: Effects of Varying Incentive and Criterion of Success on Task Performance. *Psychophysiology*, 13(5):378–385.

Lara, C. A., Flores, J., Mitre-Hernandez, H., and Perez, H. (2018). Induction of Emotional States in Educational Video Games through a Fuzzy Control System. *IEEE Transactions on Affective Computing*, 3045(c):1–12.

Larkin, K. T., Manuck, S. B., and Kasprowicz, A. L. (1989). Heart rate feedback-assisted reduction in cardiovascular reactivity to a videogame challenge. *The Psychological Record*, 39(3):365–371.

Larkin, K. T., Manuck, S. B., and Kasprowicz, A. L. (1990). The effect of feedback-assisted reduction in heart rate reactivity on videogame performance. *Biofeedback and Self-Regulation*, 15(4):285–303.

Larkin, K. T., Zayfert, C., Veltum, L. G., and Abel, J. L. (1992). Effects of feedback and contingent reinforcement in reducing heart rate response to stress. *Journal of Psychophysiology*, 6(2):119–130.

Larsen, J. T. J., Berntson, G. G., Poehlmann, K. M., Ito, T. A., Cacioppo, J. T., and Berntson, L. G. (2008). The psychophysiology of emotion. In Lewis, R., Haviland-Jones, J. M., and Barrett, L. F., editors, *The handbook of emotions*, chapter Ch.11, pages 180–195. Guilford Press, New York, 3rd ed. edition.

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

Léger, P.-m., Charland, P., Feldstein, H. D., Robert, J., Babin, G., and Lyle, D. (2011). Business simulation training in information technology education: guidelines for new approaches in IT training. *Journal of Information Technology Education*, 10(1):39–53.

Lehrer, P. and Vaschillo, E. (2004). Heart Rate Variability Biofeedback: A New Tool for Improving Autonomic Homeostasis and Treating Emotional and Psychosomatic Diseases (Special Address in 2002). *Japanese Journal of Biofeedback Research*, 30:7–16.

Lehrer, P. M., Vaschillo, E., and Vaschillo, B. (2000). Resonant frequency biofeedback training to increase cardiac variability: rationale and manual for training. *Applied psychophysiology and biofeedback*, 25(3):177–191.

Leiner, D., Fahr, A., and Früh, H. (2012). EDA Positive Change: A Simple Algorithm for Electrodermal Activity to Measure General Audience Arousal During Media Exposure. *Communication Methods and Measures*, 6(4):237–250.

Leng, H., Lin, Y., and Zanzi, L. (2007). An experimental study on physiological parameters toward driver emotion recognition. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 237–246. Springer.

Leslie, A. M. (1994). ToMM, ToBY, and Agency: Core architecture and domain specificity. In Hirschfeld, L. A. and Gelman, S. A., editors, *Mapping the Mind: Domain specificity in cognition and culture*, chapter 5, pages 119–147. Cambridge University Press.

Levenson, R. (1992). Autonomic nervous system differences among emotions. *Psychological science*, 3(1):23–27.

Licht, C. (2008). Association between major depressive disorder and heart rate variability in the Netherlands Study of Depression and Anxiety (NESDA). *Archives of General Psychiatry*, 65(12):1358–1367.

Licklider, J. (1960). Man-computer symbiosis. *Human Factors in Electronics, IRE Transactions on*, 22(140):4–11.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Lindley, C., Nacke, L., and Sennersten, C. (2008). Dissecting play-investigating the cognitive and emotional motivations and affects of computer gameplay. In *13th International Conference on Computer Games (CGames 2008)*, pages 3–5. Wolverhampton, UK.

Lindley, C. a. and Sennersten, C. C. (2008). Game Play Schemas: From Player Analysis to Adaptive Game Mechanics. *International Journal of Computer Games Technology*, 2008:1–7.

Liu, C., Conn, K., Sarkar, N., and Stone, W. (2008). Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Transactions on Robotics*, 24(4):883–896.

Liu, D., Li, X., and Santhanam, R. (2013). Digital Games and Beyond: What Happens When Players Compete? *MIS Quarterly*, 37(1):111–124.

Lo, A. W. and Repin, D. V. (2002). The Psychophysiology of Real-Time Financial Risk Processing. *Journal of Cognitive Neuroscience*, 14(3):323–339.

Lo, A. W., Repin, D. V., and Steenbarger, B. N. (2005). Fear and Greed in Financial Markets : A Clinical Study of Day-Traders. *American Economic Review*, 95(2):352–359.

Loewenstein, G. (1996). Out of Control: Visceral Influence on Behavior. *Organizational Behavior and Human Decision Processes*, 65(3):272–292.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2):426–432.

Lombard, M., Ditton, T., and Crane, D. (2000). Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Presence 2000: The Third International Workshop on Presence*, Netherlands.

Lotte, F. and Congedo, M. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2):R1–R13.

Lowenstein, O. (1955). Pupillary Reflex Shapes and Topical Clinical Diagnosis. *BMC neurology*, 5(9):631–646.

Lozano, L. M., García-Cueto, E., and Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2):73–79.

Lucas, E. (1893). *Récréations mathématiques*. Gauthier-Villars.

Lucey, B. M. and Dowling, M. (2005). The role of feelings in investor decision-making. *Journal of Economic Surveys*, 19(2):211–237.

Malcuit, G. and Beaudry, J. (1980). Voluntary heart rate lowering following a cardiovascular arousing task. *Biological psychology*, 10(3):201–210.

Malliani, A., Pagani, M., Lombardi, F., and Cerutti, S. (1991). Cardiovascular neural regulation explored in the frequency domain. *Circulation*, 84(2):482–492.

Malmstrom, E. J., Opton Jr., E., and Lazarus, R. S. (1965). Heart rate measurement and the correlation of indices of arousal. *Psychosomatic Medicine*, 27(6):546–556.

Malone, T. (1980). What Makes Things Fun to Learn? Heuristics for designing instructional computer games. In *Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*, pages 162–169. ACM.

Mandryk, R. L. (2008). Physiological measures for game evaluation. In *Game usability: Advice from the experts for advancing the player experience*, chapter 7, pages 207–235. Morgan Kaufmann Pub.

Martin, L. and Delgado, M. (2011). The influence of emotion regulation on decision-making under risk. *Journal of cognitive neuroscience*, 23(9):2569–2581.

McAllister, G., Mirza-Babaei, P., and Avent, J. (2013). Improving Gameplay with Game Metrics and Player Metrics. In *Game Analytics*, pages 621–638. Springer.

McCanne, T. R. (1983). Changes in autonomic responding to stress after practice at controlling heart rate. *Biofeedback and Self-Regulation*, 8(1):9–24.

McCaul, K. D., Solomon, S., and Holmes, D. S. (1979). Effects of paced respiration and expectations on physiological and psychological responses to threat. *Journal of Personality and Social Psychology*, 37(4):564–571.

McCraty, R., Atkinson, M., and Tiller, W. (1995). The effects of emotions on short-term power spectrum analysis of heart rate variability. *The American Journal of Cardiology*, 76(14):1089–1093.

McCraty, R., Atkinson, M., Tomasino, D., and Bradley, R. T. (2009). The coherent heart: Heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order. *Integral Review*, 5(2):10–115.

Mezzacappa, E., Kindlon, D., Saul, J. P., and Earls, F. (1998). Executive and motivational control of performance task behavior, and autonomic heart-rate regulation in children: Physiologic validation of two-factor solution inhibitory control. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39(4):525–531.

Michael, D. and Chen, S. (2005). *Serious games: Games that educate, train, and inform.* Muska and Lipman/Premier-Trade.

Millán, J. D. R., Renkens, F., Mouriño, J., and Gerstner, W. (2004). Noninvasive brain-actuated control of a mobile robot by human EEG. *IEEE Transactions on Biomedical Engineering*, 51(6):1026–1033.

Miller, A. (2015). Games Centered Approaches in Teaching Children & Adolescents : Systematic Review of Associated Student Outcomes. *Journal of Teaching in Physical Education*, 34(1):36–58.

Miller, G. a. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81–97.

Mitchell, A. and Savill-Smith, C. (2004). The use of computer and video games for learning - a review of the literature. Technical report, Learning and Skills Development Agency.

Moravec, H. (2000). *Robot: Mere machine to transcendent mind.* Oxford University Press.

Moraveji, N., Olson, B., Nguyen, T., Saadat, M., Khalighi, Y., Pea, R., and Heer, J. (2011). Peripheral paced respiration: influencing user physiology during information work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 423–428. ACM.

Moresi, S., Adam, J. J., Rijcken, J., Van Gerven, P. W. M., Kuipers, H., and Jolles, J. (2008). Pupil dilation in response preparation. *International Journal of Psychophysiology*, 67(2):124–30.

Morrison, A. (2001). The interpretation of intrusions in psychosis: an integrative cognitive approach to hallucinations and delusions. *Behavioural and Cognitive Psychotherapy*, 29(3):257–276.

Muller, M. and Ellis, D. (2011). Signal processing for music analysis. *IEEE Journal on Selected Topics in Signal Processing*, 5(6):1088–1110.

Nacke, L. (2009). *Affective ludology: Scientific measurement of user experience in interactive entertainment.* PhD thesis, Blekinge Institute of Technology.

Nacke, L. and Lindley, C. (2008). Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, Future Play '08, pages 81–88, New York, NY, USA. ACM.

Nacke, L., Lindley, C., and Stellmach, S. (2008). Log Who's Playing: Psychophysiological Game Analysis Made Easy through Event Logging. In Markopoulos, P., de Ruyter, B., IJsselsteijn, W., and Rowland, D., editors, *Fun and Games*, volume 5294 of *Lecture Notes in Computer Science*, pages 150–157. Springer Berlin / Heidelberg.

Nacke, L., Schild, J., and Niesenhaus, J. (2010). Gameplay experience testing with playability and usability surveys–an experimental pilot study. In *Proceedings of the Fun and Games 2010 Workshop, NHTV Expertise Series*, volume 10.

Nacke, L. E. (2013). An Introduction to Physiological Player Metrics for Evaluating Games. In *Game Analytics*, pages 585–619. Springer.

Nacke, L. E., Drachen, A., Kuikkaniemi, K., and Kort, Y. a. W. D. (2009). Playability and Player Experience Research. In *Proceedings of DiGRA*, pages 1–11. London.

Nacke, L. E., Kalyn, M., Lough, C., and Mandryk, R. L. (2011). Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 103–112, New York, NY, USA. ACM.

Nacke, L. E. and Lindley, C. A. (2010). Affective ludology, flow and immersion in a first–person shooter. *The Journal of the Canadian Game Studies Association*, 3(5):1–21.

Naqvi, N., Shiv, B., and Bechara, A. (2006). The Role of Emotion in Decision Making: A Cognitive Neuroscience Perspective. *Current Directions in Psychological Science*, 15(5):260–264.

Nijholt, A., Bos, D. P.-O., and Reuderink, B. (2009). Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Computing*, 1(2):85–94.

Norman, D. (2003). *Emotional design: Why we love (or hate) everyday things*. Basic Books, 1st edition.

Novemsky, N. and Kahneman, D. (2005). The boundaries of loss aversion. *Journal of Marketing Research*, 42(2):119–128.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., and Patton, M. W. (2011). Embodied Conversational Agent-Based Kiosk for Automated Interviewing. *Journal of Management Information Systems*, 28(1):17–48.

O'Donohue, W. T. and Fisher, J. E. (2009). *General principles and empirically supported techniques of cognitive behavior therapy*. John Wiley & Sons.

Oh, W. (2002). C2C versus B2C: A comparison of the winner's curse in two types of electronic auctions. *International Journal of Electronic Commerce*, 6(4):115–138.

Omar, H. M. and Jaafar, A. (2010). Conceptual framework for a heuristics based methodology for interface evaluation of educational games. *Computer and Information Science*, 3(2):211–219.

Ortiz De Guinea, A., Titah, R., and Léger, P. M. (2013). Measure for Measure: A two study multi-trait multi-method investigation of construct validity in IS research. *Computers in Human Behavior*, 29(3):833–844.

Ouwerkerk, M. (2011). Unobtrusive emotions sensing in daily life. In *Sensing Emotions*, pages 21–40. Springer.

Ouwerkerk, M., Pasveer, F., and Langereis, G. (2008). Unobtrusive sensing of psychophysiological parameters. In *Probing Experience*, pages 163–193. Springer Netherlands.

Pagani, M., Lombardi, F., and Guzzetti, S. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation research*, 59(2):178–193.

Papillo, J. and Shapiro, D. (1990). The cardiovascular system. In Cacioppo, J. and Tassinary, L., editors, *Principles of Psychophysiology: Physical, Social,* *and Inferential Elements.*, chapter 9, pages 456–512. Cambridge University Press, Cambridge.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, 4(1-2):5–20.

Parnandi, A., Ahmed, B., Eva, S., and Gutierrez-Osuna, R. (2013). Chill-Out: Relaxation training through respiratory biofeedback in a mobile casual game. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, volume 130, pages 252–260. Springer Verlag.

Parnandi, A. and Gutierrez-Osuna, R. (2017). Visual Biofeedback and Game Adaptation in Relaxation Skill Transfer. *IEEE Transactions on Affective Computing*, 3045(c):1–15.

Parnandi, A. and Gutierrez-Osuna, R. (2018). Partial reinforcement in game biofeedback for relaxation training. In *IEEE Transactions on Affective Computing*, pages 1–14. IEEE.

Partala, T. and Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1-2):185–198.

Pearson, H. (2007). The good, the bad and the ugly. *Nature*, 447(7141):138–140.

Peck, E. M., Afergan, D., Yuksel, B. F., Lalooses, F., Jacob, R. J. K., Bekele, E., and Sarkar, N. (2014). Psychophysiological Feedback for Adaptive Human–Robot Interaction (HRI). In Fairclough, S. H. and Gilleade, K., editors, *Advances in Physiological Computing*, chapter 7, pages 117–139. Springer-Verlag London, London.

Peffer, G., Cederholm, H., Clough, G., and Jercic, P. (2010). Evaluating Games Designed to Improve Financial Capability. In *Proceedings of 9th European Conference on e-Learning (ECEL'2010)*, pages 808–811.

Peffers, K., Tuunanen, T., Rothenberger, M. a., and Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77.

Pehlivanoglu, D., Jain, S., Ariel, R., and Verhaeghen, P. (2014). The ties to unbind: Age-related differences in feature (un)binding in working memory for emotional faces. *Frontiers in Psychology*, 5:1–13.

Pereira, A., Prada, R., and Paiva, A. (2012). Socially present board game opponents. In *Advances in Computer Entertainment*, pages 101–116. Springer Berlin / Heidelberg.

Perski, A. and Engel, B. T. (1980). The role of behavioral conditioning in the cardiovascular adjustment to exercise. *Biofeedback and Self-Regulation*, 5(1):91–104.

Peterson, R. (2007). Affect and financial decision-making: how neuroscience can inform market participants. *The Journal of Behavioral Finance*, 8(2):70–78.

Pfeifer, R., Iida, F., and Lungarella, M. (2014). Cognition from the bottom up: On biological inspiration, body morphology, and soft materials. *Trends in Cognitive Sciences*, 18(8):404–413.

Pfeifer, R., Lungarella, M., and Iida, F. (2012). The challenges ahead for bio-inspired'soft'robotics. *Communications of the ACM*, 55(11):76–87.

Phelps, E. a. (2006). Emotion and cognition: insights from studies of the human amygdala. *Annual review of psychology*, 57:27–53.

Picard, R. (2000). *Affective Computing*. The MIT Press, 1st edition.

Picard, R., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191.

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. In *Theories of emotion*, chapter 8, pages 3–33. Academic press.

Pope, A. T., Stephens, C. L., and Gilleade, K. (2014). Biocybernetic Adaptation as Biofeedback Training Method. In Fairclough, S. H. and Gilleade, K., editors, *Advances in Physiological Computing*, chapter 5th, pages 91–115. Springer-Verlag London, London, UK.

Porges, S. W. (1995). Cardiac vagal tone: a physiological index of stress. *Neuroscience and biobehavioral reviews*, 19(2):225–233.

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, 74(2):116–143.

Posner, J., Russell, J. a., and Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734.

Powers, A. and Kiesler, S. (2007). Comparing a computer agent with a humanoid robot. In *Proceedings of 2004 ACM/IEEE International Conference on Human-robot Interaction*, pages 145–152, Washington D.C.

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., and Aguilar, M. (2008). The pupil dilation response to visual detection. In *Proceedings of SPIE-IS and T Electronic Imaging*, volume 6806, pages 6806 – 6811. Spie.

Rainville, P., Bechara, A., Naqvi, N., and Damasio, A. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1):5–18.

Rani, P., Liu, C., Sarkar, N., and Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis and Applications*, 9(1):58–69.

Rani, P., Sarkar, N., and Adams, J. (2007). Anxiety-based affective communication for implicit human–machine interaction. *Advanced Engineering Informatics*, 21(3):323–334.

Rasmusen, E. (2006). *Games and information: An introduction to Game Theory.* Wiley-Blackwell, 4th edition.

Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., and Kivikangas, M. (2006). Spatial Presence and Emotions during Video Game Playing: Does It Matter with Whom You Play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Recarte, M. A. and Nunes, L. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology Applied*, 9(2):119–133.

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

Reeves, B., Wise, K., Maldonado, H., Kogure, K., Shinozawa, K., and Naya, F. (2003). Robots Versus On-Screen Agents: Effects on Social and Emotional Responses. In *The Conference on Human Factors in Computing Systems (CHI 2003)*, Fort Lauderdale. ACM.

Reeves, J. L. and Shapiro, D. (1982). Heart Rate Biofeedback and Cold Pressor Pain. *Psychophysiology*, 19(4):393–403.

Rehnmark, F. and Bluethmann, W. (2005). Robonaut: the'Short List'of technology hurdles. *Computer*, 38(1):28–37.

Reitan, R. M. and Wolfson, D. (1996). Theoretical, Methodological, and Validational Bases of the Halstead-Reitan Neuropsychological Test Battery. In Goldstein, G., Beers, S. R., and Hersen, M., editors, *Comprehensive Handbook of Psychological Assessment: Intellectual and Neuropsychological Assessment, Volume 1*, chapter 8, pages 105–131. John Wiley & Sons, Inc.

Reivich, K. J., Seligman, M. E. P., and McBride, S. (2011). Master Resilience Training in the U.S. Army. *American Psychologist*, 66(1):25–34.

Repetto, C., Gaggioli, A., Pallavicini, F., Cipresso, P., Raspelli, S., and Riva, G. (2013). Virtual reality and mobile phones in the treatment of generalized anxiety disorders (GAD): A phase-2 clinical trial. *Personal and Ubiquitous Computing*, 17(2):253–260.

Rieber, L. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development*, 44(2):43–58.

Riedl, R., Banker, R., and Benbasat, I. (2009). On the foundations of NeuroIS: reflections on the Gmunden Retreat 2009. *Communications of the Association forInformation Systems*, 27(15):243–264.

Riedl, R., Hubert, M., and Kenning, P. (2010). Are there neural gender differences in online trust? An fMRI study on the perceived trustworthiness of eBay offers. *Mis Quarterly*, 34(2):397–428.

Riedl, R., Kindermann, H., Auinger, A., and Javor, A. (2012). Technostress from a Neurobiological Perspective. *Business and Information Systems Engineering*, 4(2):61–69.

Riedl, R., Mohr, P., Kenning, P., Davis, F., and Heekeren, H. (2011). Trusting humans and avatars: behavioral and neural evidence. In *Proc 32nd international conference on information systems (ICIS 2011)*, pages 1–23, Shanghai. AIS Electronic Library (AISeL).

Riley, D. M. and Furedy, J. J. (1981). Effects of Instructions and Contingency of Reinforcement on the Operant Conditioning of Human Phasic Heart Rate Change. *Psychophysiology*, 18(1):75–81.

Ritterfeld, U., Cody, M., and Vorderer, P. (2009). *Serious games: Mechanisms and effects*. Taylor and Francis, 1st edition.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*. John Wiley and Sons, 2nd edition.

Rolls, E. T. and Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Progress in neurobiology*, 86(3):216–244.

Ros, R., Baroni, I., and Demiris, Y. (2014). Adaptive human–robot interaction in sensorimotor task instruction: From human to robot dance tutors. *Robotics and Autonomous Systems*, 62(6):707–720.

Rose, R. D., Buckey, J. C., Zbozinek, T. D., Motivala, S. J., Glenn, D. E., Cartreine, J. A., and Craske, M. G. (2013). A randomized controlled trial of a self-guided, multimedia, stress management and resilience training program. *Behaviour Research and Therapy*, 51(2):106–112.

Rowe, D., Sibert, J., and Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 480–487. ACM.

Rubin, R., Rubin, A., Graham, E., Perse, E., and Seibold, D. (2010). *Communication research measures II: A sourcebook*. Routledge, New York, 1st edition.

Russell, J. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178.

Russell, J. a. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172.

Sanfey, A., Rilling, J., and Aronson, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626):1755–1758.

Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th international conference on Human-robot interaction - HRI '11*, pages 305–312. ACM.

Saul, J. (1990). Beat-to-beat variations of heart rate reflect modulation of cardiac autonomic outflow. *Physiology*, 5(1):32–37.

Saul, J. and Albrecht, P. (1987). Analysis of long term heart rate variability: methods, 1/f scaling and implications. *Computers in cardiology*, 14:419–422.

Sayers, B. (1973). Analysis of heart rate variability. *Ergonomics*, 16(1):17–32.

Schaaff, K., Mueller, L., Kirst, M., and Heuer, S. (2011). Improving financial decision making using psychophysiological sensor data. In *51st Annual Meeting of Society for Psychophysiological Research (SPR)*, pages 87–103. Wiley-Blackwell.

Schachter, S. and Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5):379–399.

Schaeffer, J. (1997). *One jump ahead: challenging human supremacy in checkers*. Springer- Verlag, New York, 1st edition.

Schalk, G., Brunner, P., Gerhardt, L., Bischof, H., and Wolpaw, J. (2008). Brain-computer interfaces (BCIs): Detection instead of classification. *Journal of Neuroscience Methods*, 167(1):51–62.

Scheirer, J., Fernandez, R., Klein, J., and Picard, R. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with computers*, 14(2):93–118.

Schell, J. (2008). *The Art of Game Design: A book of lenses.* CRC Press, 1st edition.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4):695–729.

Scholl, B. J. and Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8):299–309.

Schwartz, G. E. (1972). Voluntary control of human cardiovascular integration and differentiation through feedback and reward. *Science*, 175(4017):90–93.

Scott, R. W., Peters, R. D., Gillespie, W. J., Blanchard, E. B., Edmunson, E. D., and Young, L. D. (1973). The use of shaping and reinforcement in the operant acceleration and deceleration of heart rate. *Behaviour research and therapy*, 11(2):179–185.

Sennersten, C. C. and Lindley, C. A. (2009). An Investigation of Visual Attention in FPS Computer Gameplay. In *Games and Virtual Worlds for Serious Applications, 2009. VS-GAMES'09. Conference in*, pages 68–75. IEEE, IEEE.

Seo, M.-G. and Barrett, L. (2007). Being emotional during decision making—good or bad? An empirical investigation. *Academy of Management Journal*, 50(4):923–940.

Sequeira, H., Hot, P., Silvert, L., and Delplanque, S. (2009). Electrical autonomic correlates of emotion. *International Journal of Psychophysiology*, 71(1):50–56.

Sharry, J., McDermott, M., and Condron, J. (2003). Relax To Win Treating children with anxiety problems with a biofeedback video game. *Eisteach*, 2:22–26.

Shefrin, H. and Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of finance*, 40(3):777–790.

Shiv, B., Loewenstein, G., Bechara, A., Damasio, H., and Damasio, A. R. (2005). Investment behavior and the negative side of emotion. *Psychological science*, 16(6):435–9.

Shoukry, L., Sturm, C., and Galal-Edeen, G. H. (2015). Pre-MEGa: A Proposed Framework for the Design and Evaluation of Preschoolers' Mobile Educational Games. In *Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering*, pages 385–390. Springer.

Sidner, C., Lee, C., Kidd, C., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164.

Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N. (2004). Where to look: a study of human-robot engagement. In Funchal, editor, *Proceedings of the 9th international conference on Intelligent user interfaces*, pages 78–84, Madeira, Portugal. ACM New York, NY.

Sisbot, E. and Alami, R. (2005). Navigation in the presence of humans. In *Proc. 5th International Conference on Humanoid Robots (IEEE-RAS)*, pages 181–188. IEEE.

Sliwinski, J., Katsikitis, M., and Jones, C. M. (2017). A Review of Interactive Technologies as Support Tools for the Cultivation of Mindfulness. *Mindfulness*, 8(5):1150–1159.

Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, 177(3):1333–1352.

Smith, K. and Dickhaut, J. (2005). Economics and emotion: Institutions matter. *Games and Economic Behavior*, 52(2):316–335.

Sohaib, A. T., Qureshi, S., Hagelbäck, J., Hilborn, O., and Jerčić, P. (2013). Evaluating classifiers for emotion recognition using EEG. In *Foundations of Augmented Cognition. AC 2013. Lecture Notes in Computer Science*, volume 8027, pages 492–501. Springer Berlin Heidelberg.

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., and Phelps, E. a. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13):5035–5040.

Sosnowski, T., Krzywosz-Rynkiewicz, B., and Roguska, J. (2004). Program running versus problem solving: mental task effect on tonic heart rate. *Psychophysiology*, 41(3):467–75.

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational researcher*, 35(8):19–29.

Squire, K. and Jenkins, H. (2003). Harnessing the power of games in education. *Insight*, 3(1):5–33.

Squire, L. R., Berg, D., Bloom, F., du Lac, S., and Ghosh, A. (20012). *Fundamental neuroscience*, volume 1. Academic press, 4th edition.

Stanners, R., Coulter, M., and Sweet, A. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion*, 3(4):319—-340.

Steinhauer, S. R., Siegle, G. J., Condray, R., and Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, 52(1):77–86.

Stern, R., Ray, W., and Quigley, K. (2001). *Psychophysiological recording*. Oxford University Press, 2nd edition.

Su, C.-H. (2016). The effects of students' motivation, cognitive load and learning anxiety in gamification software engineering education: a structural equation modeling study. *Multimedia Tools and Applications*, 75(16):10013–10036.

Susi, T. and Johannesson, M. (2007). Serious games: An overview. Technical report, School of Humanities and Informatics, University of Skövde.

Sütterlin, S., Herbert, C., Schmitt, M., Kübler, A., and Vögele, C. (2011). Frames, decisions, and cardiac-autonomic control. *Social neuroscience*, 6(2):169–177.

Syrdal, D. and Dautenhahn, K. (2006). 'Doing the right thing wrong'-Personality and tolerance to uncomfortable robot approaches. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, pages 183–188. IEEE.

Tan, G., Fink, B., Dao, T. K., Hebert, R., Farmer, L. S., Sanders, A., Pastorek, N., and Gevirtz, R. (2009). Associations among pain, PTSD, mTBI, and heart rate variability in veterans of operation enduring and Iraqi Freedom: A pilot study. *Pain Medicine*, 10(7):1237–1245.

Tapus, A. (2013). Coping with Stress Using Social Robots as Emotion-Oriented Tool: Potential Factors Discovered from Stress Game Experiment. In *International Conference on Social Robotics*, pages 160–169. Springer International Publishing.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. a., Ranta-Aho, P. O., and Karjalainen, P. a. (2014). Kubios HRV–heart rate variability analysis software. *Computer methods and programs in biomedicine*, 113(1):210–20.

Thayer, J. F., Hansen, A. L., Saus-Rose, E., and Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, 37(2):141–153.

Thrun, S. and Bennewitz, M. (1999). MINERVA: A second-generation museum tour-guide robot. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 3, pages 1999–2005. IEEE.

Tira-Thompson, E. and Halelamien, N. (2004). Tekkotsu: Cognitive Robotics on the Sony AIBO. In *Abstract for poster at International Conference on Cognitive Modeling*, Pittsburgh, PA. ICCM.

Vaitl, D. (1996). Interoception. *Biological psychology*, 42(1-2):1–27.

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE review*, 41(2):1–16.

van Erp, J., Thurlings, M., Brouwer, A.-M., and Werkhoven, P. (2011). BCIs in Multimodal Interaction and Multitask Environments: Theoretical Issues and Initial Guidelines. In Stephanidis, C., editor, *Universal Access in Human-Computer Interaction. Users Diversity*, volume 6766 of *Lecture Notes in Computer Science*, pages 610–619. Springer Berlin / Heidelberg.

Varvogli, L. and Darviri, C. (2011). Stress management techniques: Evidence-based procedures that reduce stress and promote health. *Health Science Journal*, 5(2):74–89.

Vaschillo, E. G., Vaschillo, B., and Lehrer, P. M. (2006). Characteristics of resonance in heart rate variability stimulated by biofeedback. *Applied Psychophysiology Biofeedback*, 31(2):129–142.

Villani, D., Carissoli, C., Triberti, S., Marchetti, A., Gilli, G., and Riva, G. (2018). Videogames for Emotion Regulation: A Systematic Review. *Games for Health Journal*, 7(2):85–99.

vom Brocke, J., Riedl, R., and Léger, P.-M. (2012). Application Strategies for Neuroscience in Information Systems Design Science Research. *Journal of Computer Information Systems*, 53(3):1–25.

Vossel, G. and Zimmer, H. (1990). Psychometric properties of non-specific electrodermal response frequency for a sample of male students. *International Journal of Psychophysiology*, 41(5):593–609.

Vrana, S., Cuthbert, B., and Lang, P. (1986). Fear imagery and text processing. *Psychophysiology*, 23(3):247–253.

Vyzas, E. and Picard, R. (1998). Affective pattern classification. Technical report, The Media Lab, MIT, Orlando, Florida.

Wainer, J., Dautenhahn, K., Robins, B., and Amirabdollahian, F. (2013). A Pilot Study with a Novel Setup for Collaborative Play of the Humanoid Robot KASPAR with Children with Autism. *International Journal of Social Robotics*, 6(1):45–65.

Wallace, J. C., Edwards, B. D., Shull, A., and Finch, D. M. (2009). Examining the Consequences in the Tendency to Suppress and Reappraise Emotions on Task-Related Job Performance. *Human Performance*, 22(1):23–43.

Walters, M. (2005). The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, pages 347–352.

Walters, M., Woods, S., Koay, K., and Dautenhahn, K. (2005). Practical and methodological challenges in designing and conducting human-robot interaction studies. In *Proceedings of the Artificial Intelligence and Simulation of Behaviour (AISB)*. AISB.

Wang, H.-m. and Huang, S.-c. (2014). Musical Rhythms Affect Heart Rate Variability. *Advances in Electrical Engineering*, 2014(851796):1–14.

Wang, Q., Sourina, O., and Nguyen, M. K. (2010). EEG-Based "Serious" Games Design for Medical Applications. In *2010 International Conference on Cyberworlds*, pages 270–276. IEEE.

Wang, Q., Sourina, O., and Nguyen, M. K. (2011). Fractal dimension based neurofeedback in serious games. *The Visual Computer*, 27(4):299–309.

Watson, D., Wiese, D., Vaidya, J., and Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76(5):820–838.

Weber, M. and Camerer, C. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior and Organization*, 33(2):167–184.

Williams, L. and Bargh, J. (2009). The unconscious regulation of emotion: Nonconscious reappraisal goals modulate emotional reactivity. *Emotion*, 9(6):847–854.

Wilson, G. F., Russell, C. a., Monnin, J. W., Estepp, J. R., and Christensen, J. C. (2010). How Does Day-to-Day Variability in Psychophysiological Data Affect Classifier Accuracy? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(3):264–268.

Winton, W. M., Putnam, L. E., and Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology*, 20(3):195–216.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international* *conference on evaluation and assessment in software engineering*, pages 1–38. ACM.

Wohlin, C., Höst, M., and Henningsson, K. (2003). Empirical research methods in software engineering. In *Empirical Methods and Studies in Software Engineering. Lecture Notes in Computer Science*, volume 2765, pages 7–23. Springer Berlin Heidelberg.

Wolf, K., Mass, R., Ingenbleek, T., Kiefer, F., Naber, D., and Wiedemann, K. (2005). The facial pattern of disgust, appetence, excited joy and relaxed joy: an improved facial EMG study. *Scandinavian journal of psychology*, 46(5):403–409.

Xin, M. and Sharlin, E. (2006). Exploring human-robot interaction through telepresence board games. In *Advances in Artificial Reality and Tele-Existence*, pages 249–261, Heidelberg. Springer Berlin Heidelberg.

Xin, M. and Sharlin, E. (2007). Playing games with robots-A method for Evaluating Human-Robot Interaction. In Sarkar, N., editor, *Human-Robot Interaction*, chapter 26, page 522. Itech Education and Publishing, Vienna.

Xu, Y., Liu, G., Hao, M., Wen, W., and Huang, X. (2010). Analysis of affective ECG signals toward emotion recognition. *Journal of Electronics (China)*, 27(1):8–14.

Yamato, J., Brooks, R., Shinozawa, K., and Naya, F. (2003). Human-Robot Dynamic Social Interaction. *NTT Technical Review*, 1(6):37–43.

Yannakakis, G. N., Martinez, H. P., and Garbarino, M. (2016). Psychophysiology in Games. *Emotion in Games*, pages 119–137.

Yerkes, R. M. and Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, 18(5):459–482.

Zafar, M. A., Ahmed, B., Al-Rihawi, R., and Gutierrez-Osuna, R. (2018). Gaming away stress: Using biofeedback games to learn paced breathing. *IEEE Transactions on Affective Computing*, 3045(c):1–14.

Zafar, M. A., Ahmed, B., and Gutierrez-osuna, R. (2017). Playing with and without Biofeedback. In *Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference on.*, pages 1–7. IEEE.

Zagal, J., Mateas, M., Fernandez-Vara, C., Hochhalter, B., and Lichti, N. (2007). Towards an Ontological Language for Game Analysis. In Jenson, J. and de Castell, S., editors, *Worlds in play: International perspectives on digital games research*, chapter 2, pages 21–34. Springer, 1st edition.

Zammitto, V., Seif El-Nasr, M., and Newton, P. (2010). Exploring quantitative methods for evaluating sports games. In *Proc. CHI 2010 Workshop on Brain, Body and Bytes*.

Zhai, J. and Barreto, A. (2006a). Stress detection in computer users based on digital signal processing of noninvasive physiological variables. *IEEE Engineering in Medicine and Biology Society. Conference*, 1(Ic):1355–8.

Zhai, J. and Barreto, A. (2006b). Stress detection in computer users through non-invasive monitoring of physiological signals. *Biomedical Sciences Instrumentation*, 42:495–500.

Zoghbi, S., Parker, C., Croft, E., and der Loos, H. V. (2010). Enhancing Collaborative Human-Robot Interaction Through Physiological-Signal Based Communication. In *Proceedings of Workshop on multimodal human–robot interfaces at IEEE International Conference on Robotics and Automation (ICRA 2010)*. IEEE.

ABSTRACT

Emotions are thought to be one of the key factors that critically influence human decision-making. Emotion-regulation can help to mitigate emotion-related decision biases and eventually lead to a better decision performance. Serious games emerged as a new angle introducing technological methods to practicing emotion-regulation, where meaningful biofeedback information communicates player's affective states to a series of informed gameplay choices. These findings motivate the notion that in the decision context of serious games, one would benefit from awareness and regulation of such emerging emotions.

This thesis explores the design and evaluation methods for creating serious games where emotion-regulation can be practiced using physiological biofeedback measures. Furthermore, it investigates emotions and the effect of emotion-regulation on decision performance in serious games. Using the psychophysiological methods in the design of such games, emotions and their underlying neural mechanism have been explored.

The results showed the benefits of practicing emotion-regulation in serious games, where decision-making performance was increased for the individuals who down-regulated high levels of arousal while having an experience of positive valence. Moreover, it increased also for the individuals who received the necessary biofeedback information. The results also suggested that emotion-regulation strategies (i.e., cognitive reappraisal) are highly dependent on the serious game context. Therefore, the reappraisal strategy was shown to benefit the decision-making tasks investigated in this thesis. The results further suggested that using psychophysiological methods in emotionally arousing serious games, the interplay between sympathetic and parasympathetic pathways could be mapped through the underlying emotions which activate those two pathways. Following this conjecture, the results identified the optimal arousal level for increased performance of an individual on a decision-making task, by carefully balancing the activation of those two pathways. The investigations also validated these findings in the collaborative serious game context, where the robot collaborators were found to elicit diverse affect in their human partners, influencing performance on a decision-making task. Furthermore, the evidence suggested that arousal is equally or more important than valence for the decision-making performance, but once optimal arousal has been reached, a further increase in performance may be achieved by regulating valence. Furthermore, the results showed that serious games designed in this thesis elicited high physiological arousal and positive valence. This makes them suitable as research platforms for the investigation of how these emotions influence the activation of sympathetic and parasympathetic pathways and influence performance on a decision-making task.

Taking these findings into consideration, the serious games designed in this thesis allowed for the training of cognitive reappraisal emotion-regulation strategy on the decision-making tasks. This thesis suggests that using evaluated design and development methods, it is possible to design and develop serious games that provide a helpful environment where individuals could practice emotion-regulation through raising awareness of emotions, and subsequently improve their decision-making performance.

ISSN: 1653-2090

ISBN: 978-91-7295-370-3

2019:06