Chapter 11

Discussion

The collection of articles presented in this thesis all contributed to the research questions posed in Chapter 1. In this final chapter, the overall results of those works are discussed and related to the overarching theme of this thesis; 'Measuring and modeling negative emotions for virtual training'. This is done in Section 11.1 by revisiting the four research questions and their connections to the various chapters. Section 11.2 goes on to give an overview of the software that played a large part in the research done for this thesis and discusses both past experiences and future possibilities. Considering future possibilities, Section 11.3 mentions some ethical considerations when working with (extreme) negative emotions and the potential risks thereof, which should be kept in mind when continuing this research, either as detailed in Section 11.4 or in another direction.

11.1 Research questions

Table 11.1 presents an overview per chapter of its relevance for each of the research questions posed in Chapter 1. It can be clearly seen that the first part dealt more with the the psychological side of this work, while the second part covered the computational modeling aspect. There is a distinct separation visible between the two parts, which will be discussed in Section 11.4, but here each research question will be addressed separately with respect to the contributions each chapter made in answering it.

RQ 1 *Is it possible to evoke negative emotions using virtual stimuli?*

Four out of five chapters in part I are concerned with experiments that tried to evoke some sort of negative emotion using various stimuli. Although the goals of the various studies differ slightly, within each of these experiments participants were confronted with an emotion inducing stimulus and asked about their subjective experience. As such, they contribute to this first research question and the paragraphs below re-iterate the relevant findings per chapter before discussing them as a whole.

Chapter 2. Scary movies, difficult computer games and unfair evaluations were used to induce stress. Participants did indicate they experienced more

		RQ 1	RQ 2	RQ 3	RQ 4
Part I	Chapter 2	✓	✓		
	Chapter 3		✓		
	Chapter 4	✓	✓		
	Chapter 5	✓		✓	
	Chapter 6	✓		✓	
Part II	Chapter 7				✓
	Chapter 8				✓
	Chapter 9				✓
	Chapter 10				✓

Table 11.1: Research questions and its relevance per chapter

stress under those conditions that were specifically designed to do so. Interestingly, even though some care should be taken when interpreting these results due to slight differences in the questionnaires, a comparison of the three different experiments showed clear variations to what extent the stimulus was able to elicit certain mental states. For example, video did not only prove to be capable of inducing a subjective feeling of stress, but gave way to the condition indicated to be the least stressful. Games on the other hand, even those chosen for their low level of difficulty, always resulted in relatively high ratings on stress. And then again, injustice, compared to the other two stimuli, seemed to be experienced as being less stressful.

Chapter 4. The goal of this research was to compare real-life stimuli with a virtual counterpart. In order to do so, an aggressive outburst from a real human was compared with that of a virtual avatar which was animated using recordings of that same person. From the questionnaire, the only difference found between the two conditions was that participants listened more carefully to the real human. No differences were found in believability of the outburst, the scariness of it or even if they felt personally addressed. Although physiological measurements told a slightly different story (discussed below), these were promising results.

Chapter 5. Although the focus of this chapter was more on RQ 3, it is also relevant here as pictures were used to induce negative emotions and participants had to indicate how strongly they experienced those feelings. The images used were taken from a standardized set composed specifically to induce particular feelings. As such, the participants in this experiment did indicate to experience those negative emotions.

Chapter 6. Again, the aim of this evaluation study was more towards investigating the potential to use virtual training than it was on evoking emotions. Nevertheless, the participants did experience a virtual simulation of a situation which would be designated as stressful in real life. A subjective evaluation of this simulation resulted in positive scores in most areas, except for the emotional aspect where unfortunately the scores were mainly negative.

Discussion. In these chapters, a large variety of stimuli was used to induce negative emotions: images, movies, games, injustice, an aggressive outburst and aggressive passengers. With the exception of the last one, each stimulus resulted in a clear subjective experience of a negative emotion. The virtual passengers used in Chapter 6 did show aggressive behavior, but the focus in this study was mainly on decision making. Therefore, the expressed aggression was not as intense as for example shown by the virtual avatar in Chapter 4, which could well be the cause that the participants in this study did not feel scared or threatened. For all the other stimuli, participants indicated that they felt stressed, scared or negatively aroused to some degree. And it is exactly this slight difference in semantics that makes it difficult to compare the various stimuli. Even though in Chapter 2 such a comparison has been attempted with experiments where more or less the same questions were used to obtain subjective ratings of stress, putting the subjective ratings of these chapters next to each other is troublesome. Not only are there differences in the targeted emotion, people often have difficulty in rating their own experience accurately as discussed below. Nevertheless, these chapters do show people can experience negative emotions when confronted with virtual stimuli and that subjectively some types of stimuli can be more stressful than others. Video for example can be used effectively both for relaxation and strong emotional experiences. Games on the other hand are always experienced to be stressful to some extent, but are also the most consistent across participants in generating feelings of stress. Virtual avatars can elicit negative emotions similar to those a real human can evoke, however the aggression shown by such virtual characters might have to be relatively intense before such an emotional response occurs.

RQ 2 *How can the psychological experience of negative emotions using virtual stimuli be measured accurately?*

Chapters 2 through 4 use physiological sensors and attempt to find a relation between subjective ratings of emotion and more objective measurements of heart rate, skin conductance or brain waves.

Chapter 2. As explained above, three different stimuli were used to elicit an experience of stress. In none of these conditions was that feeling reflected in the heart rate measurements, although it did appear to respond to the active participation required in two of the three experiments. Skin conductance on the other hand clearly differentiated between a baseline condition and a stressful one. Although there were differences as well between the three stimuli with respect to the skin conductance, here no satisfactory explanation for that effect could be found.

Chapter 3. During the video experiment of Chapter 2, not just skin conductance and heart rate were measured, but brain activity was measured as well, using consumer-grade EEG hardware. Although it was nearly impossible to differentiate between the various conditions on the basis of averages, some specific moments did result in clearly visible peaks in various frequency bands. Manually, explanations could be found for the largest of those peaks by looking at the corresponding video material. However, for some peaks such a relation could not be found and the film contained scenes similar to those resulting in large peaks where (almost) no change in brain activity could be noticed.

Chapter 4. No changes in heart rate were found while skin conductance clearly responded to the aggressive outburst used in this chapter to elicit a stress response, which is in line with the findings in Chapter 2. As discussed above, no subjective differences were found between the experience of a virtual avatar getting angry and a real human doing the same. However, in the virtual condition skin conductance was lower throughout and the increase during the aggressive outburst was less pronounced. And even though heart rate did not respond to the aggressive outburst, in the real condition heart rate was higher on average for the entire experiment.

Discussion. While RQ 1 focused on the subjective experience of stress, the chapters regarding RQ 2 covered the physiological response. Throughout the experiments in Chapters 2 and 4 skin conductance levels increased consistently when negative emotions were being elicited. On the other hand, heart rate did not show a clear response in any of the chapters, which can be explained by the inherent difficulties using heart rate to this end as discussed in Section 1.2.3. Interestingly, Chapter 4 clearly shows the importance of using physiological data and not only relying on self-report measurements. While no differences were found in how participants rated their emotional experience between the two conditions, both skin conductance and heart rate did reveal large differences. Moreover, the increase in skin conductance when faced with a real human becoming angry (± 4µS) seems to be of a different order of magnitude compared to any of the reactions when faced with virtual stimuli (between 1 and 2µS on average). This shows the difficulty people can have in rating their emotional experience and strengthens the need for more objective measurements. To add to the arsenal of tools for obtaining such measurements, Chapter 3 explored the possibilities of using consumer-grade EEG devices. Even though the hardware is inferior to the costly and complex professional EEG sensors used in for example neurological research, results do show its potential for measuring stress in run-time and warrant more extensive research in this area.

RQ 3 *Is virtual training useful in learning to cope with negative emotions?*

Two chapters in this thesis consider the learning effects of a virtual training. While Chapter 5 focuses on emotion regulation, Chapter 6 is directed at improving decision making in a stressful situation.

Chapter 5. Three groups of participants had to subjectively rate their emotional response after seeing a picture. At a later time point, two of those groups received a task to either regulate their emotion when viewing those pictures or identify whether the image depicted something positive or negative. The third group did nothing. When revisiting the pictures, it turned out that the emotional experience can either be weakened (by the first task) or strengthened (by the second task). Those effects persisted even after six months and applied to other images for the group that learned to regulate their emotions while viewing such images as well.

Chapter 6. Professionals in public transport underwent a weekly training for four weeks in which they had to de-escalate threatening situations with aggressive passengers using a virtual simulation. Various analyses showed little to no learning effects, even though subjectively participants indicated they felt they were better able to cope with those situations and considered such a training a useful addition to what is currently offered using role-play. This lack of improvement could be related to the absence of a difference in performance between the less and more experienced employees, indicating it is difficult to improve on this skill even with more experience. Another explanation for the lack of improvement could be that additional feedback was necessary on top of the outcome of the scenario, to improve the performance of trainees.

Discussion. While training was successful in Chapter 5, no improvement of performance was found in Chapter 6, but it might be difficult to compare the two. In Chapter 5, the group that successfully learned to cope with scary or horrific pictures were given a specific training to regulate their emotions. Within the other experiment participants ‘just’ practiced in a virtual environment and received no specific pointers or feedback. Thus, the first experiment focused on emotion regulation with a specific training compared to the latter targeting decision making with straightforward practice sessions. There are various possibilities for improving this training software. For example, it was found there was no significant difference in performance of experienced professionals compared to newer employees, which makes it questionable whether a trainee can even improve over time without specific feedback. A second important aspect for improvement is adaptivity. For now, each participant underwent the same scenarios, even though the pre-test already revealed differences in skill. And as a last remark, the absence of an appropriate emotional experience during the scenario could have limited presence and immersion, which subsequently might affect learning as well.

RQ 4 *How can computational models (of emotion) be used for virtual training?*

The second part of this thesis is concerned with methods to use computational models (of emotion) to improve virtual training. Chapters 7 through 10 each investigated a different approach to do so.

Chapter 7. A computational model of emotion regulation has been developed based on various existing theories. Simulations of the resulting model show expected patterns for emotion regulation and an attempt was made to reproduce experimental data using the model. Even though there is room for improvement, estimations of emotion levels clearly proved more accurate using the computational model of emotion regulation.

Chapter 8. In this chapter the possibilities of using a computational model (in this case of emotion regulation) as a tool in the design of a virtual training are explored. By simulating human behavior using this model, the performance of a proposed system can be formally analyzed. With this approach it is possible to refrain from testing with human participants in the early stages of development, which can be costly and time consuming.

Chapter 9. The model described in this chapter is related to social groups and encompasses a method of using computational models to generate behavior, taking into account the social groups, personal preferences, and the situational context. By defining so-called role prototypes, the behavior of virtual avatars can be generated as a function of the social groups they belong to. Using this approach, agents can show conflicted behavior when the social groups they relate to require different actions in a particular context.

Chapter 10. Based on theories regarding aggression de-escalation, a model to select a suitable level of difficulty is discussed. First, a distinction is made between two different types of aggression. Second, three different learning goals are identified. The model defines a method for trainees to progress from the easier learning goals to more difficult ones for both types of aggression separately dependent on the performance.

Discussion. At first glance, these chapters might seem fairly unrelated. However, looking back at Figure 1.1 (p.5) where the architecture used in the project STRESS is depicted, the various topics can be related to the different components of that design as was briefly touched upon in the introduction as well. Chapter 7 describes a computational model which can be used as part of the 'affective model' determining the mental state of a trainee. Chapter 8 builds on this work and develops analysis and support models as a tool for the design of a virtual training. However, the analysis model can be used just as well for analyzing a trainee's mental state ('affective model'), while the support model is suitable for reasoning about the best possible way for trainees to control their emotions ('feedback determination module'). Although the model described in Chapter 9 is related to cultural behavior, the method of using such computational models to generate behavior can be applied to emotional models as well. By incorporating these models in virtual avatars, the resulting behavior of those characters will be conform the theories underlying the model. Characters exposed to similar stimuli, but using different emotion regulation strategies, might exhibit different behavior and react differently as well. As such, the displayed behavior in the 'virtual reality environment' will be more believable as well. As for Chapter 10, there is a clear link with the 'scenario development module' and the method for determining the difficulty adaptively, although developed for aggression de-escalation, can easily be adapted to work for other learning goals as well. And while this thesis does not contain any research into the 'decision making model', such work has been done as part of the project and is described in for example Bosse & Provoost (2014) or Bosse & Schnitfink (2015).

11.2 Software

During the years of working on this thesis, a number of different software packages and programming techniques were used. In this section, some experiences of various tools are provided. To start with, a couple of notes on programming the psychological experiments are given. Methods used for using or integrating various physiological sensors are mentioned in the subsequent section. A separate section is devoted to the training software used in Chapter 6, named InterAct, which has been developed especially for the project STRESS. To implement such models as are described in Part II, a toolbox for Matlab has been created during the course of this research that should facilitate easy integration with other soft- and hardware.

11.2.1 Psychological experiments

For the development of the experiments in Chapters 2 to 5 an application named PsychoPy has been used.[1] In essence, this is a set of libraries for the programming language Python combined with a graphical user interface to aid inexperienced programmers with creating psychological experiments. Although this interface was not used, the various libraries made it easy to implement experiments with accurate timing and correct logging of results. Moreover, as it is built upon a commonly used programming language, some of the physiological measurements could be easily integrated as discussed hereafter.

11.2.2 Physiological sensors

Physiological sensors from two manufacturers were used in this research. Both the heart rate measurements and skin conductance levels were recorded using hardware from PLUX.[2] For the EEG measurements used in Chapter 3 a Myndplay Brainband was used.[3] These two pieces of hardware are discussed separately in the following two paragraphs.

Biosignals PLUX. A large variety of sensors can be used with this physiological device, among those required for measuring heart rate and skin conductance. Via a bluetooth connection, this data is transferred to a computer on which these measurements can be visualized and recorded using software distributed with the hardware. There are however also programming interfaces available for a variety of languages, making it possible to integrate physiological data in for example the experiments developed in PsychoPy. Using this method, the recorded data consists of the raw measurements and needs to be processed in order to obtain the desired values in the appropriate units. To this end, a couple of scripts were created in Matlab such that after each experiment the raw data could be transformed into the desired format. Although this approach was workable for these experiments, a more extensive implementation is required when the raw data needs to be converted in real-time, which may prove a necessity in future work.

[1] Available from http://www.psychopy.org (accessed 21-8-2015).
[2] A newer version is available via http://biosignalsplux.com/ (accessed 21-8-2015).
[3] Available from http://myndplay.com/ (21-8-2015).

Myndplay Brainband. This low-cost EEG device is based on the same chipset of the better known company Neurosky.[4] Although it is shipped with its own software, the options that exist for using a Neurosky device can be applied as well. Having said that, for Chapter 3 the provided software was used to start recording brain signals manually and were matched to the corresponding video using timestamps after the experiment. However, to integrate this data in other applications directly, a number of possibilities exist. At the time of writing, various efforts have been made to integrate Myndplay Brainband data in other applications. For example, using the so-called ThinkGear Communications Driver provided by Neurosky, the device has been successfully connected to the InterAct software described below. With the help of a shared library, it is possible to connect to the device using a large variety of programming languages and access both the raw and processed data at any moment. Thus, when such a connection has been established, using the brain data is relatively straightforward.

11.2.3 InterAct

During the first years of the project STRESS, an external company (IC3D media)[5] started developing the virtual environment for simulating 1 on 1 conversations to be used for training purposes within that project. This software has been used in Chapter 6 to evaluate a training for public transport personnel. Even though performance of participants did not seem to improve during these four weeks, the majority enjoyed using the software and would like to continue doing so. As the training used in that chapter did not use the full potential of the software package, the next two paragraphs highlight two features that will be beneficial for future research to improve on the current training.

Faceshift. Whilst it is often possible to freely create a dialogue structure, InterAct allows for adding custom animations (including speech) as well. Using a commercial software package named Faceshift,[6] facial expressions can be captured fairly easy and the corresponding speech is recorded as well. Virtual characters can then be animated using those recordings as is done in Chapter 4, but is also possible for the characters within InterAct. For those characters, a large library of arm and hand gestures is also available to add to the scenario as desired. Using this approach, the same software can be used as well for training police agents (as is being done at the moment of writing) or in any other domain without the need to involve professional animators.

[4] http://neurosky.com/ (accessed 21-8-2015).
[5] http://ic3dmedia.com/ (accessed 21-8-2015).
[6] http://www.faceshift.com/ (accessed 21-8-2015).

Network node. A second feature that is not yet used in the research included here is the so-called network node. This adds the possibility to connect via TCP/IP to InterAct and incorporate external software within the scenario. When such a connection is established, the external software can read and write variables used in the scenario as well as display messages within the training environment. This allows the developer to incorporate models as described in part II, without the need to involve the software developers of InterAct.

11.2.4 L2-matlab

The computational models described in part II have been implemented in various ways using Excel, LEADSTO, Java or Python. Excel for example is suitable for implementing quantitative models (Chapter 7), but when qualitative concepts play a role as well LEADSTO has the advantage (Chapter 8). However, both of these tools are difficult to connect with external software such as for example InterAct or the software that is used by the physiological sensors. Java (Chapter 9) or Python (Chapter 10) are suitable to that end, but do require quite some effort to implement and simulate models in the first place. Although Matlab has not been used for any of the models in this thesis, it does work well for implementing numerical models. Moreover, Matlab also includes an extensive programming language and using that we have developed an extension of Matlab. This extension, or toolbox, named l2-matlab,[7] has the same benefits as LEADSTO with regards to easy implementation of models using both quantitative and qualitative concepts, plus it has the flexibility of a broadly used programming language which opens up the possibilities to connect with other software using various methods.

Using a couple of text files with a .l2 extension the structure of a model is defined. Matlab functions are then used to implement the dynamics of the model. Furthermore, scenarios and parameters can be defined in .l2 files as well. Simulating the model and plotting the results can be done using the functions provided with the toolbox. There is an option as well to perform parameter tuning using an evolutionary algorithm on any of the parameters used in the model. Appendix A contains a more extensive description on how to use this toolbox and includes a number of examples as well. A large number of students have already used the software to implement a wide variety of models during relevant courses. In future research, this software could potentially be used to implement computational models, include physiological data and connect with for example software such as InterAct all in one.

[7] http://www.few.vu.nl/~jmn300/l2-matlab/ (accessed 21-8-2015).

11.3 Ethical considerations

There is an assumption underlying the STRESS project that a trainee needs to experience negative emotions on some level to be able to effectively learn how to cope with such stressful situations in daily life. Therefore, much of the research carried out for this thesis involved evoking those negative emotions using virtual stimuli. However, as was pointed out in Section 1.2.2 (p.4), there are risks when exposed to negative emotions with an extremely high intensity. With technology progressing fast, it is important to consider at one point a virtual environment starts to pose similar risks as the 'real' world.

For a large part of this research, pictures and video were used to evoke negative emotions. Before those experiments took place, participants were explicitly asked if they normally could cope with watching a scary or horror movie. On that basis, some participants were not allowed to join the experiment. For the others, risks were low as the stimuli were presented on a regular computer screen and this hardly differs from simply watching a scary movie. For the larger experiments using the virtual environment approval was given by an ethical committee. Again, the risks were low as the virtual environment was presented on a computer screen, participants could easily abort the training and the aggression displayed was of a short duration.

However, a logical progression of this research would be to take the environment from the computer screen to a more immersive medium such as a cave automatic virtual environment (CAVE) or a head-mounted display. This increased immersion will surely add to the experienced emotion and caution should be taken when exposing participants to negative emotions. With increasing realism, there is a growing risk of virtual experiences posing same psychological risks as their real-life counterparts. Occasionally, researchers identify this risk in their work (i.e. Kinateder et al., 2014), but more often experiments are performed without mentioning such considerations. For example, Reiners et al. (2014) goes as far as to recreate the fear of dying in order to raise awareness for health and safety issues, but in doing so neglects to raise the problem of the risk involved with the virtual reality used in their own experiment.

Currently, virtual stimuli are not able to recreate the exact experience of the real world, as shown in Chapter 4 or related work discussed in Section 1.2.5. In line with the Media Equation theory (Reeves & Nass, 1996), people do tend to treat virtual stimuli as if they were real, which is supported by both subjective, behavioral or physiological responses in the relevant chapters or any of the referenced experiments throughout this thesis. At the same time, there appears to be a structural difference in the intensity in which the emotion is experienced.

So far only a handful of articles have compared real and virtual experiences. Borst & Gelder (2015) performed an extensive review of the literature that concerns how virtual characters are perceived compared to humans. They found that people react in a similar way to virtual characters as they do to actual human beings, even on an emotional and neurological level. However, most research that is mentioned is limited to the perception of the avatar or only looks for similar response patterns without explicitly comparing a virtual with a real condition. As an illustration of this last point, consider the following paper mentioned by Borst & Gelder. Slater et al. (2006) replicated the famous experiment of Milgram (1963) using a virtual avatar. They found that participants 'tended to respond to the situation at the subjective, behavioral and physiological levels as if it were real'. This cannot accurately be compared to the original (real-life) experiment though, as such measurements were not precisely recorded then.

In 2007, Raij et al. did compare the virtual and real conditions directly with a group of medical students that had to elicit information from either a real or a virtual patient. Even though the same information was gathered from the virtual patient, participants showed less engagement which was attributed to the limited expressive behavior of the avatar. In a more recent study, Hays et al. (2012) performed a similar experiment in which trainees had to practice interpersonal and counseling skills, comparing role-play with a human to that with a virtual avatar. Both self-report measurements and physiological data were obtained to compare the two conditions, and here no significant differences between them were found. These results notwithstanding, the nature of the task given in this study is in itself not very emotional and differences could be obscured by a minimal emotional response to begin with, or perhaps differences only start to occur with more extreme stimuli.

An important question to be answered is why there is such a difference between emotional responses. Hartmann (2012) discus the reason for this disparity based on the guilt felt when shooting a virtual avatar and makes a distinction between feeling as if something is real opposed to knowing it is not. This knowledge that something is not real might then be related to the differences that are found between the emotional responses. Whether this is indeed the case needs more research, but is important for both the risks and effects of virtual training.

If participants would intuitively know a virtual environment is not real, and thus have less intense emotional responses, this would probably limit the risks of negative emotions in a virtual environment. If this is the case, does this then also limit the learning potential of such an environment? Also, if participants currently are able to regulate their emotions rather well based on the knowledge that the experience is virtual, the question would be how this is achieved, if it is sufficient to limit the potential negative effects and again what the effects are on learning. Alternatively, it might be that currently this knowledge exists due to the virtual environment not being immersive or realistic enough. In that case, the knowledge that the experience is not real might fade away completely with technology progressing and a virtual environment will start to pose real-life risks. In any case, there is a need for more research in this area in order to continue this work in a safe and sensible way.

11.4 Future work

The previous section described some considerations and directions for future research from an ethical perspective. There are however numerous different ways to continue with the work described in this thesis that would be worthwhile. This section highlights various possibilities for future work.

RQ 1 and RQ 2 were related to evoking and measuring negative emotions. In this work these emotions were elicited using different kinds of virtual stimuli all displayed on a regular computer screen. We found that the level of interaction already played a part in the emotions experienced (though not subjectively) and it would be interesting to investigate in what other ways the chosen medium relates to the intensity of the emotions. In Riva et al. (2007) a link is made between presence and emotion experienced in a virtual environment. Furthermore, IJsselsteijn et al. (2001) show an effect of the type of medium on presence. Thus, it might be concluded that by changing the medium the emotional response will change accordingly, but this is not necessarily true. Although Reeves et al. (1999) found a clear effect of screen size on arousal, research comparing for example head-mounted displays with regular screens on the emotional experience are hard to find. Closely related is the effect of a specific medium on learning. For example, Moreno & Mayer (2002) found an effect of the medium used on presence, but not on performance. Similarly, in an experiment conducted by Fassbender et al. (2012), learning did not improve using a more immersive display. Thus, even though changing the method of presentation might affect the emotional experience, it is important to consider the effects on learning separately.

The second part of this thesis covered methods to use computational models for virtual training. However, none of these models were implemented in an application to be tested with real users and thus this is left for future work. Although Chapter 6 did cover the evaluation of software for training aggression de-escalation, no computational models were incorporated in that case. No improvement in skills was found, even though participants did indicate otherwise and would like to continue using such methods for training. This indicates it is important to investigate if incorporating computational models would actually result in a better performance of trainees, not just in a subjective idea of a better performance. If computational models could help improve virtual trainings, physiological measurements could provide us with better models to begin with. In Chapter 7 an attempt is made to estimate subjective ratings of an emotional response using a computational model of emotion regulation with moderate success. Such an estimate could be improved by using physiological measurements as additional information, which in turn can improve the virtual training in which those models are incorporated. In other words, exciting times have arrived and work can be started on integrating the two parts!

References

Borst, A. W. d., & Gelder, B. d. (2015). Is it the real deal? perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Frontiers in Psychology*, *6*(576).

Bosse, T., & Provoost, S. (2014). Towards aggression de-escalation training with virtual agents: A computational model. In P. Zaphiris, & A. Ioannou (Eds.) *Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration*, vol. 8524 of *Lecture Notes in Computer Science*, (pp. 375–387). Springer International Publishing.

Bosse, T., & Schnitfink, K. (2015). The effect of simulated threat on task performance during emotion recognition. In D. Harris (Ed.) *Engineering Psychology and Cognitive Ergonomics*, vol. 9174 of *Lecture Notes in Computer Science*, (pp. 107–116). Springer International Publishing.

Fassbender, E., Richards, D., Bilgin, A., Thompson, W. F., & Heiden, W. (2012). Virschool: The effect of background music and immersive display systems on memory for facts learned in an educational virtual environment. *Computers & Education*, *58*(1), 490–500.

Hartmann, T. (2012). Moral disengagement during exposure to media violence. In R. Tamborini (Ed.) *Media and the moral mind*, (pp. 109–131). Routledge.

Hays, M. J., Campbell, J. C., Trimmer, M. A., Poore, J. C., Webb, A. K., & King, T. K. (2012). Can role-play with virtual humans teach interpersonal skills? *Proceedings of the 34th interservice/industry training, simulation, and education conference*.

IJsselsteijn, W., Ridder, H. d., Freeman, J., Avons, S. E., Bouwhuis, D., et al. (2001). Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence. *Presence*, *10*(3), 298–311.

Kinateder, M., Ronchi, E., Nilsson, D., Kobes, M., Muller, M., Pauli, P., & Muhlberger, A. (2014). Virtual reality for fire evacuation research. In *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*, (pp. 313–321). IEEE.

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, *67*(4), 371–378.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of educational psychology*, *94*(3), 598–610.

Raij, A. B., Johnsen, K., Dickerson, R. F., Lok, B. C., Cohen, M. S., Duerson, M., Pauly, R. R., Stevens, A. O., Wagner, P., & Lind, D. S. (2007). Comparing interpersonal interactions with a virtual human to those with a real human. *Visualization and Computer Graphics, IEEE Transactions on*, *13*(3), 443–457.

Reeves, B., Lang, A., Kim, E. Y., & Tatar, D. (1999). The effects of screen size and message content on attention and arousal. *Media Psychology*, *1*(1), 49–67.

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

Reiners, T., Teräs, H., Chang, V., Wood, L. C., Gregory, S., Gibson, D., Petter, N., & Teräs, M. (2014). Authentic, immersive, and emotional experience in virtual learning environments: The fear of dying as an important learning experience in a simulation. In *Transformative, innovative and engaging. Proceedings of the 23rd Annual Teaching Learning Forum*. The University of Western Australia.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, *10*(1), 45–56.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A virtual reprise of the stanley milgram obedience experiments. *PloS one*, *1*(1), e39.