Exploring subjective and physiological stress responses to virtual stimuli

Employees in many different jobs have to deal with (extreme) negative emotions on a regular basis. Virtual training could be a suitable method for learning to cope with those situations if it proves possible to evoke similar emotions in such a set-up. In this research, three experiments to induce stress using different virtual stimuli are described, i.e. video, games and injustice. Each of these stimuli can be used to increase the subjective experience of stress, which is reflected in skin conductance levels as well, though not in heart rate measurements. However, by comparing the results of each experiment, heart rate does seem clearly linked to the type of stimulus, with significantly lower results in the video experiment, while there is no clear pattern in the skin conductance measurements. It is therefore important to know what such physiological responses might indicate, irrelevant of the corresponding emotion induced, in order to extract the emotional experience from those physiological measurements.

This chapter is based on:
Cusveller, J., Gerritsen, C., & Man, J. de (2014). Evoking and Measuring Arousal in Game Settings. In *Games for Training, Education, Health and Sports* (pp. 165-174). Springer International Publishing.*
Gerritsen, C., Man, J. de, & Meij, J. van der (2013). Physiological and subjective response to injustice: the effects of unjust evaluations on physiological responses and subjective experiences. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on* (Vol. 3, pp. 179-182). IEEE.*
Bosse, T., Gerritsen, C., Man, J. de, & Stam, M. (2014). Inducing anxiety through video material. In *HCI International 2014-Posters Extended Abstracts* (pp. 301-306). Springer International Publishing.*

* Authors ordered alphabetically and contributed equally to this work.

2.1 Introduction

For human beings, the ability to cope with negative emotions such as fear, stress and injustice is important to live a pleasant life. For professionals in domains such as the police, military and public transport, the specific ability to cope with fear (or anxiety) (Ohman, 2000) is presumably even more crucial. Employees in these domains generally have a higher probability of being confronted with fear-inducing stimuli like aggressive individuals, gun fights, or human casualties. And since the extreme negative emotions experienced in such situations are known to impair cognitive processes like attention and decision making (Loewenstein & Lerner, 2003; Ozel, 2001), professionals in these domains benefit greatly from effective emotion regulation skills. At the same time, these professionals are often treated unjust, which can result in stress and agitated behavior (Berger et al., 1998; Fowles, 1982). Besides these feelings of injustice potentially having long lasting negative consequences (Berg, 2011), the generated negative emotions decrease the quality of task performance even further.

For these reasons, much attention is dedicated to developing appropriate training methods for employees in those high risk jobs, to learn to cope with extreme circumstances. Such training often uses role-play, where the roles are played by co-students or professional actors. However, an important drawback of these types of training is that they are very costly, both in terms of money and time. As an alternative, training based on Virtual Reality (VR) currently receives much attention (Bouchard et al., 2011). The main goal of the STRESS project[1] is to develop an adaptive VR-based environment to train professionals how to cope with extreme negative emotions. Trainees will be placed in a virtual scenario in which they have to make difficult decisions while negative emotions are induced. By measuring certain physiological states, the system will be able to assess their emotional state at run-time as well as provide adequate support. This could be both in terms of offering suggestions on how to improve their emotion regulation and decision making, or by adapting the scenario at run-time.

Despite this promising prospect, the effectiveness of such a system crucially depends on its ability to evoke the desired level of anxiety in the trainee. Therefore, one of the research questions addressed in the STRESS project is to what extent computer generated stimuli trigger an emotional (stress) response. As a first step, in Bosse et al. (2013) the effect of affective pictures on emotional response was investigated. The results pointed out that a set of negatively valenced images triggered a significantly stronger emotional response than neutral or positive images. Moreover, participants that applied emotion

[1] http://stress.few.vu.nl (accessed 22-7-2015).

regulation strategies (in particular reappraisal (Gross, 2002)) experienced the images as significantly less intense when viewing them again, an effect that persisted six months later. In Section 2.2, this research as well as other relevant material is discussed in more detail.

However, in VR it are not just still images that a trainee is confronted with that cause stress and it is important to consider how different types of stimuli affect the emotional response. Therefore, in this article, a variety of stimuli are used to elicit negative emotions. First, instead of still images a video showing clips from scary movies is used in Section 2.3 to elicit a stress response. Then, Section 2.4 uses games as a stimulus, thus requiring the participant to actively participate during the experiment, while the stress is induced by added time-pressure. In Section 2.5, the task itself is very simple and does not cause any stress, but negative emotions are elicited by providing clearly incorrect feedback and adding negative consequences to that. These different approaches are compared in Section 2.6 in order to draw conclusions about the consequences this has on VR training in Section 2.7.

2.2 Background

Evoking an emotional (stress) response by different types of stimuli has been the focus of many research projects. In the literature, two main indicators of a stress response are reported, namely a change in heart rate and an increased skin conductance. Both these indicators are considered to be part of the physiological stress response (Epstein & Roupenian, 1970; Lazarus et al., 1963; Lin et al., 2011). However, some papers conclude that skin conductance is a more reliable indicator than heart rate (Goldstein et al., 1965; Khalfa et al., 2002; Speisman et al., 1964) as findings regarding change in heart rate are less consistent. For example, Craig & Lowery (1969) report that directly experienced stress-inducing stimuli may lead to an increased heart rate, while indirect (or vicarious) stimuli may cause a decreased heart rate.

Recently, experiments have been performed where stress responses are induced using virtual stimuli. Courtney et al. (2010) investigated the differences between pictures and both computer-generated images and videos of feared stimuli. They found no significant difference between the 'real' pictures and the computer-generated ones, while the computer-generated videos elicited a greater physiological response (measured using skin conductance) than both other conditions. In Blankendaal et al. (2015), the emotional response to an aggressive agent was compared to that when faced with a real aggressive human actor. Similar subjective and physiological responses were found for the agent and the actor, although the physiological response was less intense when confronted with the agent. Thus, there are similarities between the 'real' thing and the virtual, but we should consider the differences as well.

Similarly, research often compares a stressful condition with a particular baseline measurement. This baseline could either be measured after some form of relaxation or during a comparable but less stressful condition. For example, Brouwer et al. (2011) used a tour in the virtual environment where no task had to be performed as a baseline, while subsequent conditions did include a task within a similar virtual environment. Here, subjectively no differences in experienced stress were found although physiological results did show effects. In a very similar set-up, Hartanto et al. (2014) compared stress induced by a job interview in virtual reality to a neutral environment and found heart rate increased with added stress. In another experiment reported in the same article, they also found an effect based on the tone of the dialogue, comparing positive remarks with negative ones.

Interestingly, there is not much research comparing different types of baseline or stressful measurements. Below, three different experiments are explained, each of them more or less in line with the experiments mentioned above; a baseline measurement compared to a more stressful, though similar condition. The results of these three experiments will then be put side by side. Based on the differences between the experimental conditions it might be possible to find some effects of task characteristics on the physiological results of such experiments.

2.3 Video

There is still a large gap between affective images and affective VR material. Therefore, the current section investigates the effect of affective videos on emotional response. To this end, an experiment has been performed in which participants were asked to watch a variety of video clips while measuring their emotional response via physiological as well as subjective data.

2.3.1 Method

Thirty participants took part in the experiment, aged between 20 and 64 years old (with an average age of 33), of which 17 people were male and 13 female. The experiment took place in a secluded room at VU University in Amsterdam. Heart rate and skin conductance were measured using PLUX wireless biosensors.[2]

Each participant watched five different movies in sequence (Figure 2.1). After each movie, participants were asked to report the emotion they felt as well as its intensity. The possible emotions were ‘relaxed’, ‘bored’, ‘interested’,

[2] http://www.biosignalsplux.com (accessed 11-8-2015).

'excited' and 'scared', the intensity was given on a Likert-scale ranging from 0 (not at all) to 5 (very much). A short break of 30 seconds preceded the next movie. The first movie showed an empty beach for three minutes in order to get the participants in a calm state. The second movie showed a three minute clip from a nature documentary and was used to measure a baseline for both heart rate and skin conductance. The third clip was a collection of scenes from various scary/horror movies and was intended to evoke a stress response. The fourth movie was a different three minute clip from the same nature documentary as before. The final movie was a repetition of the empty beach. These last two movies were shown to see whether heart rate and skin conductance returned to their baseline values.

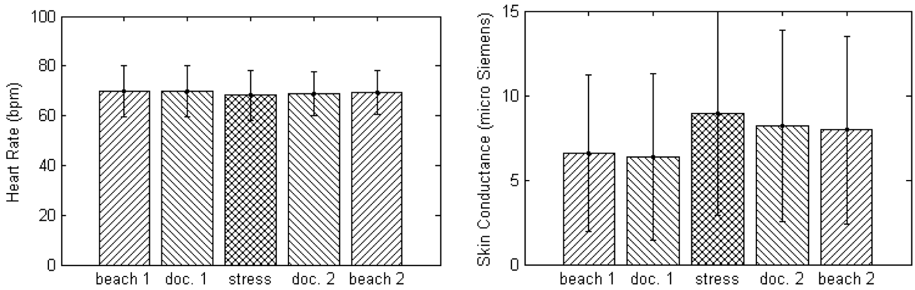
Figure 2.1: Sequence of video clips used for data collection

2.3.2 Results

Before considering any possible effects in physiological responses due to the different nature of the clips, it is first checked whether each of the movies indeed evoked the desired emotion. A pairwise comparison of the subjectively reported emotions using a Bonferroni correction showed that each clip evoked a different emotion at the $p < 0.001$ level, except for the 1st and 5th (beach) as well as 2nd and 4th movie (documentary). The beach was found to be relaxing or boring, the documentary interesting and the stressful movie exciting or scary.

To investigate whether heart rate and skin conductance of the participants during the stressful movie differed significantly from the other clips, a repeated measures ANOVA has been performed. This method is suited to test for significant differences within participants between the two experimental groups they were part of.

Figure 2.2(a) shows the average heart rate during each movie. As can be seen, there are only small differences between the clips. This is confirmed by the repeated measures ANOVA which shows no significant differences in the average heart rate ($F(4, 116) = 1.401$, $p = .238$).



(a) Heart rate (b) Skin conductance

Figure 2.2: Averages of physiological measures, including standard deviations

With regard to the skin conductance, the assumption of sphericity has been violated and a Greenhouse-Geisser correction was applied. Figure 2.2(b) shows the average values and statistical testing shows the presence of significant differences between the five clips ($F(2.251, 65.285) = 27,213$, $p < .001$). As can be seen, the average skin conductance is highest during the stressful movie, with the largest difference in comparison with the two preceding clips. A pairwise comparison reveals almost all differences to be significant at the 0.05 level. Table 2.1 shows for each pair the difference in skin conductance and its significance. While there is a significant decrease after the stressful movie, skin conductance remains significantly higher in comparison with the similar clips shown before. This is consistent with the fact that decay of skin conductance level is generally quite slow.

2.3.3 Discussion

In these experiments, a preliminary exploration has been made regarding the possibilities to induce a stress response through video material. An experiment has been performed in which participants were asked to watch five different video clips while their emotional response was measured via physiological measurements as well as questionnaires. Among the five movies, the third one was

	Doc. 1	Stressful	Doc. 2	Beach 2
Beach 1	−18.23	235.01***	161.29***	136.12**
Doc. 1		253.24***	179.53***	155.35***
Stressful			−73.71*	−97.89**
Doc. 2				−24.17
Beach 2				

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 2.1: Pairwise comparison of skin conductance differences

composed in such a way that it could be experienced as being stressful, whereas the clip prior to this movie was designed to serve as a neutral movie, to determine the baseline level of heart rate and skin conductance of the participants. The results of the measurements showed that the heart rate of the subjects did not differ significantly during the stress film compared to the other four movie clips. Instead, the skin conductance of the participants increased significantly (compared to the other four clips), and the same held for the subjective ratings. Hence, we conclude that it is possible to generate a stress response by means of video material, and that skin conductance is an effective indicator to measure this.

Although this is a promising result, it is only a first step to the accomplishment of our objectives. As mentioned earlier, our final aim is to develop a Virtual Reality-based environment for training of emotion regulation and emotional decision making skills. Hence, as a next step, we will perform similar experiments with actual video game material instead of video clips (comparable to the work of Brouwer et al. (2011)), and compare the results with our current results. This will provide further insight in the possibilities to develop intelligent VR-based training systems using physiological measurements.

2.4 Games

Serious games are games that do not have entertainment as primary purpose and are mainly used for training or education purposes (Michael & Chen, 2005). The participant should experience a high sense of presence (indicating a strong involvement of the user in the virtual environment) (Steuer, 1995), which can be achieved by dynamically adapting the (affective) content of the game to reach a desired level of arousal (Baños et al., 2004). To obtain the ‘optimal’ level of arousal it is necessary to measure arousal at run-time and to make the game adaptive to it, meaning that the game should become more challenging or exciting if the level of arousal is too low and vice versa if arousal is too high

(Nogueira et al., 2013). Following the same line of reasoning, it is important for us to keep negative emotions on an ‘optimal’ level for learning. With that in mind, this section investigates both the physiological and subjective responses when confronted with great time pressure in games.

2.4.1 Method

A computer task has been designed to measure the level of arousal in different situations while physiological measurements are registered. In the following sections, the experimental design, the participants and the hypotheses will be described.

Experimental design. In the experiment the participants had to play three (off the shelf) computer games. These games were separated by a break. A message was presented to the participants indicating that they could continue when they felt ready. The first game (Tetris) lasted approximately four minutes, the second and third game (respectively Grid16 and Guitar Maniac) two minutes each. This of course depended on the results of the player. These games were chosen based on the anticipated pressure they would bring about. The games were presented to all of the participants in the same order.

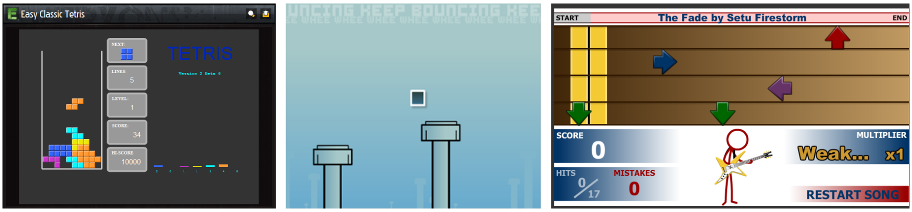
The first game that the participants played is Tetris.[3] Tetris is a well-known puzzle game in which blocks in different shapes move from top to bottom on the screen (Figure 2.3(a)). The player has to adjust the blocks so that lines will be made. A full line (from left to right) will disappear and gives the player points. If the blocks are not adjusted correctly, lines will not be made and the blocks will stack. When the blocks reach the top of the screen the game ends. Before the game started the participants received an explanation of the game and the instruction to play only the first two (easy) levels and to stop after those levels. This game was used as a baseline game. We wanted the participants to play a game which was relatively easy to do. The first two levels of Tetris are not yet very demanding and can therefore provide a baseline level in heart rate and skin conductance. The next two games were expected to be more demanding leading to a physiological response.

The second game the participants played is called Grid16.[4] In this game the participants had to play different games of skill without any explanation (see Figure 2.3(b) for a screenshot of one of those games from Grid16). For example, in these games the participants have to avoid obstacles or have to make sure an object will not fall. The player plays a game for 10 seconds and then automatically switched to another game. Meanwhile the pace increases,

[3] http://www.newgrounds.com/portal/view/235136 (accessed 11-8-2015).
[4] http://www.spelletjes.nl/spel/Grid-16.html (accessed 11-8-2015).

the games get more difficult and the system switches faster from one game to another. The system continues to switch between the games until the player is game over in all games. The participants were instructed to play the entire game twice.

The third game is called Guitar Maniac.[5] The player has to press the correct key at the right moment (Figure 2.3(c)). On the screen the participants see arrows move from right to left. The arrows are accompanied by numbers 1, 2, 3 or 4. When the arrows (and numbers) reach the gray part of the screen the participant has to press the key on the keyboard that corresponds to the number on the screen.



(a) Tetris (b) Grid16 (c) Guitar Maniac

Figure 2.3: Screenshots of the three games used for the experiment

Participants. A total of 15 participants were recruited among acquaintances of the experimenter. The age of these 9 women and 6 men ranged between 17 and 28 with an average age of 20.8. Each participant was fitted with the PLUX ECG and EDA sensors measuring heart rate and skin conductance. The heart rate was measured by placing three sensors on the skin near the heart. The skin conductance was measured by two sensors that were placed on the index- and middle finger. The participants were asked about their experience with games. Four participants indicated that they never play games, while five participants game quite frequently. The other participants game sometimes.

Subjective ratings. Before the experiment started the participants were asked to answer some questions to see whether the participant was relaxed before (s)he started the experiment and to find out how often the participant plays games. After each of the games the participants were asked if they understood the game, if they found the game difficult to play and if they found

[5] http://www.kongregate.com/games/shinki/super-crazy-guitar-maniac-deluxe-3 (accessed 11-8-2015).

the game stressful to play. The answers to these questions were rated on a 7-point scale, with 0 being indicated as never/not at all and 6 as very often/very much.

Hypotheses. The expectation was that the participants would feel stress during the faster and more complicated games. It is expected that this will show in the subjective ratings of stress with a significant increase between game 1 and games 2 and 3. For Guitar Maniac, an increase in reported difficulty of the game is expected, while for Grid16 a lowered understanding is hypothesized as well. Furthermore, it is hypothesized that this increase in stress would lead to an increased level of skin conductance in game 2 and game 3. For heart rate, stating a hypothesis is more difficult, as the increase in arousal due to stress is associated with an increased heart rate. However, as stated in Section 3, similar work did not always find any changes in heart rate.

2.4.2 Results

The following sections describe the main results of the experiment. It is divided into three sections. First the results regarding the subjective questions following each game are described. Afterward, the measurements of both skin conductance and heart rate are presented and finally correlations between the subjective and physiological results are considered.

Subjective questions. After each game, participants answered a number of questions regarding their understanding of the game, how easy the game was to play and how stressful they experienced the game to be. Figure 2.4 shows the average response of all 15 participants for each game, including the standard deviation. As can be seen, Tetris was well understood and easy to play, accompanied by a low stress score for that game. Participants had more trouble understanding Grid16, found it less easy to play and also rated the game to be more stressful. Guitar Maniac was understood by the participants, but was also found difficult to play and scored high for stress.

Statistical analysis underlines these results. A repeated one way ANOVA showed significant differences for each aspect; understanding with $F(2, 28) = 56.031$, $p < 0.001$, easy to play with $F(2, 28) = 26.793$, $p < 0.001$ and using the Greenhouse-Geisser values due to violation of the assumption of sphericity yields $F(1.44, 20.22) = 26.793$, $p < 0.001$ for stress. Post-hoc tests using the Bonferroni corrections shows all differences to be significant at the 0.05 level, except for the scores on easy to play and stress between Grid16 and Guitar Maniac. Further analysis has shown that there is a correlation between how easy each game was to play and how stressful participants rated the game

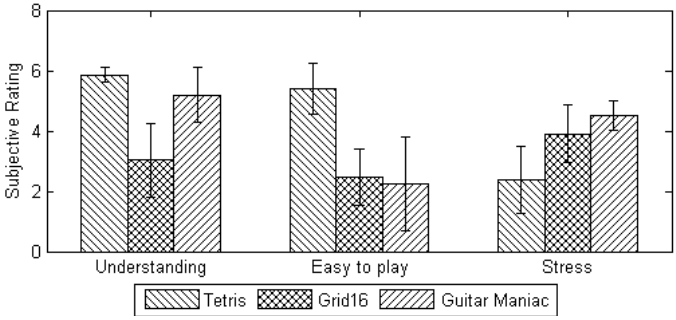


Figure 2.4: Average rating (including standard deviation) to 3 questions

($r = -0.571$, $p = 0.029$; $r = -0.651$, $p = 0.009$; $r = -0.584$, $p = 0.022$). A similar correlation was found between understanding and stress, except for guitar maniac ($r = -0.60$, $p = 0.018$; $r = -0.62$, $p = 0.013$; $r = -0.15$, $p = 0.598$).

Physiological measurements. Besides subjective questions, skin conductance and heart rate were measured to gain insight into the arousal of the participants (Figure 2.5). Repeated measures ANOVAs with Greenhouse-Geisser corrections showed statistical differences for both skin conductance ($F(1.266, 12.660) = 13.916$, $p = 0.002$) and heart rate ($F(1.437, 15.802) = 7.818$, $p = 0.008$). Post-hoc tests using a Bonferroni correction showed all differences in skin conductance to be significant at the 0.05 level, while the heart rate only significantly differs between Tetris and Grid16 as well as between Grid16 and Guitar Maniac.

Correlations between subjective and physiological measurements. Finally, it has been investigated whether there were any correlations between the subjective and physiological measures. Tetris was taken as a baseline game, providing both a subjective and physiological baseline. For each of the subjective aspects and both physiological measures, the difference between Grid16 and Tetris as well as the difference between Guitar Maniac and Tetris was calculated. Spearman's rho was calculated for each combination of subjective question and physiological measurements, but resulted in no significant correlations.

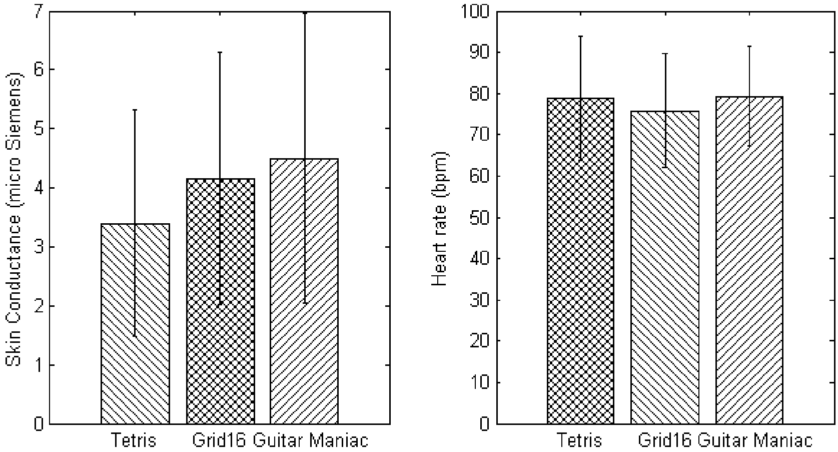


Figure 2.5: Average skin conductance (left) and heart rate (right) including standard deviation for each game

2.4.3 Discussion

In the previous section, subjective and physiological results from playing three different games were discussed. Regarding the subjective questions, playing Tetris was, as expected, less stressful for the participants. Both Grid16 and Guitar Maniac were more difficult to play, making it plausible that the easier a game is to play, the less stressful it is to the participants. For Grid16, it was found that participants had trouble understanding the game, which could have caused stress. Guitar Maniac was well understood by each participant, but nonetheless hard to play and stressful. Statistically, not understanding the game did not cause a difference in the stress experienced, but on visual inspection there is a trend towards lower stress levels when this stress is (partly) caused by incomprehension.

It was hypothesized that both the skin conductance and heart rate would increase with the more stressful games. However, this was only true for the skin conductance. The heart rate dropped significantly when playing Grid16 and no differences were found with Guitar Maniac. There is some literature showing anger to be related with a decrease in heart rate (Hassett, 1978), but it is unclear whether this applies to the situation at hand. In the end, there was no correlation found between the answers to the subjective questions and the physiological responses. It could be that the heart rate here reflected some different aspect than stress, as it did not show the expected increase over the three games. For the skin conductance however, the increase shows a similar pattern to the increase in stress experienced by the participants, but a significant correlation between the two variables could not be found. At this point in time, it is unclear whether the skin conductance is indeed not directly correlated with the stress experienced, or that a weaker correlation between the two variables is obscured by a low number of participants.

2.5 Injustice

A feeling of injustice can result in stress and agitated behavior (Berger et al., 1998; Fowles, 1982), thereby decreasing the quality of the task performance. Learning to deal with unjust situations is important for professionals who are confronted with such behavior on a regular basis. A feeling of injustice is used in this research to indicate when a person feels unjustly treated, for example when not treated according to social norms. In this section, the possibilities for evoking a feeling of injustice using computer generated stimuli are investigated.

2.5.1 Method

A computer task has been designed to invoke a feeling of injustice, while physiological measurements are registered. In the following sections, the experimental procedure and the participants will be described in more detail, followed by some hypotheses.

Experimental design. Each experiment was divided into two series. For both series, participants had to judge 30 combinations of two pictures showing human faces in sequence. In the first series, the task for group A was to pick the face that looked more happy, while in the second series the more sad looking face had to be chosen. Group B was first given the task to pick the more sad looking face and the more happy looking face in the second series. If the correct face was chosen, this was clearly indicated on a bright green screen. If the wrong photo was picked, a bright red screen indicated the error and participants were set back three combinations and had to first redo those pictures before being able to continue. Subsequently, the current progress was shown after which either the next combination of photos was shown, or an already seen combination in case of a wrong answer. This continued until the last of the 30 combinations was answered correctly or until 20 errors were made. Figure 2.6 shows this flow in graphical form.

However, in the second set, some answers were unjustly treated as being incorrect. The chance for this to happen was either 15% or, depending on

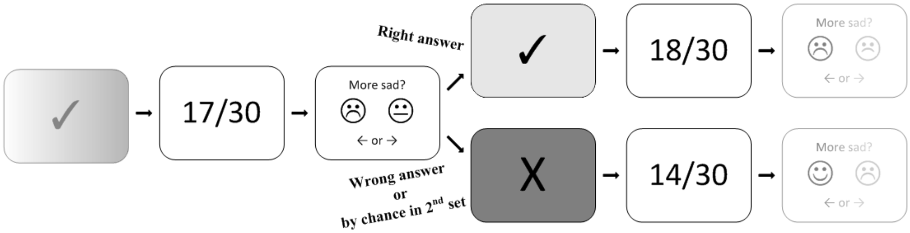


Figure 2.6: Graphical impression of experimental design. Note that in the experiment actual pictures of faces were used.

which percentage was greater, based on the progress i (between 1 and 30) as follows: $i4/8100$. Therefore, the further along people got, the greater the chance on unjust assessments with the answer to the final combination always being rejected. As the pictures were chosen such that each combination had a clear correct answer and no time limit was imposed, unjust assessments could be easily distinguished from wrong answers.

Participants. A total of 21 participants between the age of 15 and 61 with an average age of 25.5 years took part in the experiment and were randomly assigned to either group. Each participant was fitted with the PLUX ECG and EDA sensors measuring heart rate and skin conductance. The heart rate was measured by placing 3 sensors on the skin near the heart. The skin conductance was measured by two sensors that were placed on the index- and middle finger. Due to an abundance of measurement noise with a few participants some data was excluded from analysis, leaving 18 heart rate measurements (9 male and 9 female) and 17 skin conductance measurements (8 male and 9 female) of which respectively 11 and 10 belonged to group A and 7 to group B.

Subjective ratings. After each series the participants were asked to answer two questions. The answers to these questions were rated on a 6-point scale. The first question was whether the participant found the test difficult. The second question was whether the participant found the test to be stressful. These questions were included to see if the experiment actually had an effect on the subjective experience of the participant.

Hypothesis. The expectation was that the participants would feel unjustly treated when they were sure that they had given a correct answer but nevertheless received a negative evaluation and set back three places. It is expected this will show in the subjective ratings for stress with a significant increase for the second set. Furthermore, it is hypothesized that this would lead to an increased level of skin conductance in the second set of pictures and no change in heart rate, based on the experiment performed by Markovsky (1988). However, it might also be argued that an increase in heart rate should be expected. Injustice is often accompanied by emotions like anger, sadness or fear (Mikula et al., 1998) which in turn are characterized with an increase in heart rate (Ekman et al., 1983).

2.5.2 Results

Before the results of the two different series are analyzed, the subjective questions regarding task difficulty and experienced stress were used to investigate possible differences between the two groups. An unpaired t-test showed no reason to belief the order of images had any effect on the subjective experience. Therefore, in the subsequent analyses Group A and Group B are treated as being equal and only the differences between the first (just) series and the second (unjust) series are compared.

Subjective questions. Each participant was asked about their subjective experience of task difficulty and stress for both the just (1) and unjust (2) series. These results are shown in Figure 2.7. A paired t-test showed that for both the difficulty and the stressfulness differences between both series are significant ($t(21) = -5.08$, $p < 0.0001$ and $t(21) = -6.11$, $p < 0.0001$ respectively).

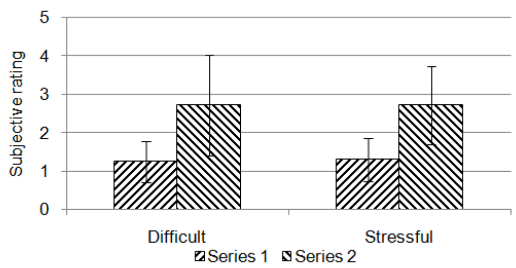
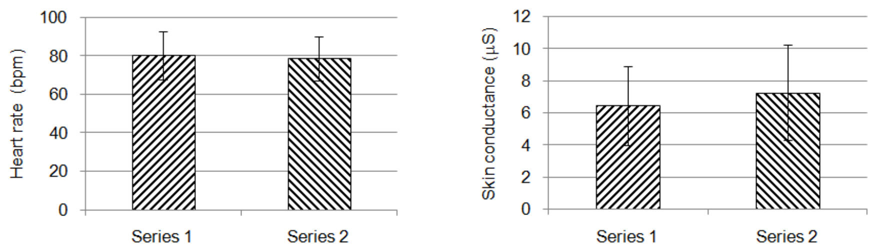


Figure 2.7: Average subjective ratings including standard deviation

Heart rate. A paired t-test was performed to test whether there is a difference in heart rate between the two sets of pictures. Although a slight decrease of 1.5 beats per minute (bpm) was found, as can be seen in Figure 2.8(a), this decrease appeared not to be significant ($t(17) = 1.81$, $p = 0.088$).

Skin conductance. Figure 2.8(b) shows the results of the skin conductance measurements. To test the significance of the 0.81 microsiemens (μS) increase between the two series, a paired t-test has been performed as well. The difference in skin conductance appeared to be significant with $t(16) = -3.52$, $p = 0.0028$. This increase was in accordance with our hypothesis.



(a) Heart rate (b) Skin conductance

Figure 2.8: Averages of physiological measures, including standard deviations

Correlations. Finally, we looked at possible correlations between the subjective questions and the physiological measurements. We calculated for each aspect the difference between the first and the second series and used these differences to calculate the Pearson correlation. Unfortunately, both the differences in heart rate and in skin conductance did not correlate with either the subjective increase in difficulty or stress. However, a significant positive correlation was found between the two subjective questions ($r = 0.569$, $p = 0.004$).

2.5.3 Discussion

The participants in the experiment were presented two series of pictures in which they had to rate the most happy or sad of two pictures. In one of the series the system deliberately judged the choice of the participant incorrectly with a chance that increased as the experiment prolonged.

As expected the level of skin conductance indeed increased during the second set of pictures. Also the heart rate seemed not to be affected by the injustice, which is in line with the experiment by Markovsky (1988). This is an important finding since it shows that the results from physiological measurements in both the real world experiment and the virtual world experiment are similar.

Unfortunately, no correlation between the physiological measurements and the subjective ratings has been found. However, we did find a correlation between the question of the task being difficult and the task being stressful which might be interesting. Initial tests showed no differences between the order of tasks (more sad or more happy), which implies that both tasks are similar. However, the correlation between the two subjective questions implies that unjust assessments not only affect the amount of stress experienced, but also the experienced task difficulty. Thus, injustice can affect the subjective experience of task difficulty, while the task itself is of a comparable difficulty. This finding underlines the importance of training professionals how to cope with unjust behavior.

2.6 Comparison

In the previous sections three experiments were described in which a stressful negative emotion is elicited using different methods. Each experiment was performed in a similar fashion, measuring heart rate, skin conductance and subjective ratings of emotional intensity. In the following sections, the results of those experiments will be compared on these aspects. Of each experiment, one condition is taken as a baseline and compared with the condition that evoked the greatest (subjective) response. This results in a comparison between a baseline and a stressful condition between the three experiments as shown in Table 2.2. In the following sections, this comparison is first made on the basis of the subjective ratings and afterward the physiological results are put side by side. These findings are discussed in the last section.

	Video	**Games**	**Injustice**
Baseline	Beach 1	Tetris	Series 1 (just)
Stressful	Stressful	Guitar Maniac	Series 2 (unjust)

Table 2.2: Conditions of each experiment as used in the comparison

2.6.1 Subjective responses

In each experiment, participants were asked about their subjective experience. Although the exact questions asked differed between the experiments, it is possible to extract a subjective rating reflecting the intensity of a negative emotion experienced. For both the games and injustice, participants were queried for the level of stress directly, but for the video both an emotion and intensity were asked for. By ordering these emotions and subsequently summing up the scores, one overall rating is calculated. For each experiment and each condition, normalized values of these scores are shown in Figure 2.9.

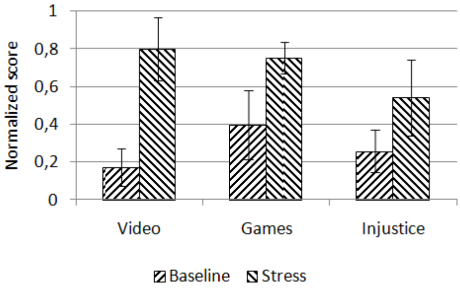


Figure 2.9: Comparison of the subjective ratings

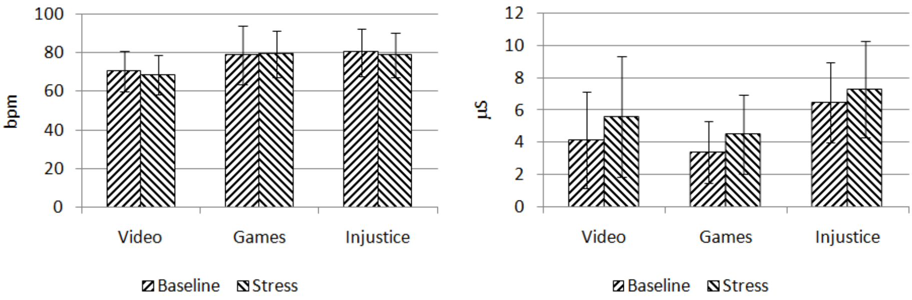
Considering the baseline measures, it is clear that for the games experiment a higher baseline value was measured than in the other two tasks (video-games $t(44) = -5.44$, $p < 0.0001$; games-injustice $t(34) = 2.83$, $p = 0.0078$). Moreover, the baseline value in the video experiment was significantly lower than that measured for injustice ($t(50) = -2.91$, $p = 0.0054$). Considering the stressful condition, both the video and games experiment obtain high ratings. Although there is no significant difference between these two values, both are significantly higher than the stressful condition of the injustice experiment (video-injustice $t(50) = 8.33$, $p < 0.0001$; games-injustice $t(34) = 3.78$, $p = 0.0006$).

2.6.2 Physiological response

In each of the experiments, both heart rate and skin conductance were measured using the same methods. Thus, it is possible to compare these values between the three different stimuli, as is shown in Figure 2.10.

Taking a closer look at the heart rate values in Figure 2.10(a), there is little difference between the baseline conditions and the stressful conditions as has been discussed above. However, it appears that during the video condition participants had a lower heart rate than in the other two experiments. These differences are statistically confirmed for the baseline measurements (video-games $t(43) = -2.20$, $p = 0.0331$; video-injustice $t(47) = -2.97$, $p = 0.0046$), as well as the stressful condition (video-games $t(41) = -2.91$, $p = 0.0057$; video-injustice $t(47) = -3.17$, $p = 0.0027$). Between games and injustice no statistically significant differences were found.

Next are the skin conductance values as shown in Figure 2.10(b). As had already become clear in the individual experiments, there are large differences in skin conductance levels between participants as visualized in the standard deviations. There are again rather large differences between the experiments as well, especially between games and injustice where both the baseline ($t(27) = -3.60$, $p = 0.0012$) and the stressful measurement ($t(26) = -2.58$, $p = 0.0160$) are significantly different. Between video and injustice there is only a difference between the baseline measurements ($t(46) = -2.75$, $p = 0.0084$), while there are no differences in either measurement between video and games.



(a) Heart rate (b) Skin conductance

Figure 2.10: Comparison of the physiological responses

As a last comparison, Figure 2.11 shows the average difference in heart rate and skin conductance between the stressful and baseline condition for each experiment. As can be seen more clearly now, on average the heart rate decreases in each experiment, but there is a very large variation between participants. As such, there were no significant differences found between the conditions in the individual experiments. Moreover, between the experiments there is also no difference to be found when comparing the change in heart rate caused by the stressful condition. Similarly, when looking at the change in skin conductance for each experiment, it is clear that the stressful condition results in a higher value. However, no significant differences can be found in the extent of that increase.

2.6.3 Discussion

Based on the subjective ratings, there are a couple of differences in the levels of stress experienced. However, it is important to note that for the video experiment there was no question directly relating to the level of stress. The comparison between games and injustice on the other hand is straightforward; both conditions in the games experiment were more stressful compared with the injustice experiment. The baseline measurement for games was based on playing Tetris, which is not yet very demanding but does involve some pressure in terms of time and task level. The baseline for injustice was a very simple

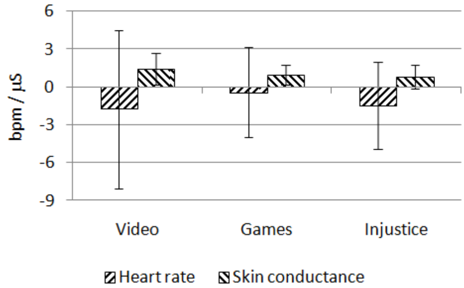


Figure 2.11: Comparison of the physiological responses

task with no time pressure, which can be considered less stressful as the results show. An interesting find for the stressful condition is the decreasing standard deviation in games, while for injustice the standard deviation increases. Thus, time pressure seems to be experienced as stressful in a similar manner between participants, while injustice was more stressful for some than others. But how does this compare to the video experiment? On the one hand we find that watching a video of a beach is the least stressful condition, which sounds reasonable. On the other hand, the stressful condition is comparable to the stressful game albeit with a greater standard deviation. Although watching a video is a passive task, the quality of movie clips is very high and it seems plausible that merely watching such a clip can be very stressful. The large standard deviation can be explained by some people not being very sensitive to such video material as they might watch such films on a regular basis. Thus, even though it cannot be ignored that these ratings are retrieved using a different method, results are in line with expectations.

Regarding the physiological measurements, there are very few differences in heart rate. The only finding here is the heart rate being lower in both conditions during the video experiment compared to the other two. The difference between those experiments is the passive nature of watching video, while for games and injustice participants had an active task throughout. Thus, even though adding stress did not result in a difference in heart rate, there seems to be an effect on the heart rate when active participation is required. In order to substantiate this claim however, a more specific experiment should be carried out.

The skin conductance measurements are difficult to compare, as there are large differences between participants. It does turn out that for games, skin conductance is lower in both conditions compared to the injustice experiment. Although not significant, the results of the video experiment seem to be in between the values for games and injustice. Thus, within the experiment, skin conductance does increase for the stressful condition, but overall the games experiment resulted in significantly lower skin conductance levels compared to the injustice experiment. However, when looking at the relative change between the baseline and stressful condition, no differences can be found between the experiments. Thus, a plausible reason for this difference in absolute values could be some external factor, for example a greater environmental temperature during the injustice experiment resulting in higher values in general. Nevertheless, more research is required to substantiate this explanation which still could reveal that there is some underlying reason for the large differences in average skin conductance levels between the three experiments.

2.7 Conclusion

In this article three separate experiments were discusses in which a form of stress was induced by a virtual stimulus. In the first experiment, video clips of scary / horror movies were successfully used as a source for stress resulting in an increased subjective experience of stress. Skin conductance measurements showed a similar increase, while heart rate did not respond to the stressful condition. For the second experiment, Tetris was compared with two other games in which time-pressure was high. Here, similar results were found; subjectively the stress experienced increased, which did show in skin conductance measurements as well but not in heart rate. For the third experiment, injustice was successfully used to induce stress. Again, skin conductance increased accordingly, while heart rate did not show any significant changes.

Taken individually, it is clear that adding a form of stress in each experiment had its effect, mostly on the subjective ratings and the skin conductance. Across experiments, it is interesting to see how the subjective ratings compare to each other. For the baseline measurements, the experiments in increasing order of stressfulness are video, injustice and games. However, for the stressful condition, injustice induced the least amount of stress while no difference in stressfulness was found between video or games. Thus, while games always induce some level of stress, videos can be both very relaxing and just as stressful as games. Thus, is important to consider your type of baseline measurement well in order to filter out the specific stress added by a scary video, time pressure in games or injustice during a simple task.

Regarding the skin conductance, there was no clear effect of the type of experiment. However, for heart rate, the video experiment resulted in significantly lower heart rates for the whole experiment compared to games and injustice. This could be due to the passive nature of watching video compared to the active role of participants in both the other experiments. For this to be confirmed, more research is required. Moreover, it would be useful to investigate what the cause is for the different skin conductance values between experiments, if there even is one. Knowing what such physiological responses might indicate, irrelevant of the corresponding emotion induced, would help in extracting the emotional experience from those physiological measurements.

Acknowledgments. This research was supported by funding from the National Initiative Brain and Cognition, coordinated by the Netherlands Organisation for Scientific Research (NWO), under grant agreement No. 056-25-013

References

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: their impact on the sense of presence. *CyberPsychology & Behavior*, *7*(6), 734–741.

Berg, J. v. d. (2011). *De politie bedreigd. Een kwalitatief onderzoek naar de gevolgen voor het welzijn van de politiefunctionaris en diens gezin.*. Master's thesis, Utrecht University.

Berger, J., Fisek, M. H., Norman, R. Z., Wagner, D. G., Berger, J., & Zelditch, M. (1998). Formation of reward expectations in status situations. *Status, power, and legitimacy*, (pp. 121–154).

Blankendaal, R., Bosse, T., Gerritsen, C., de Jong, T., & de Man, J. (2015). Are aggressive agents as scary as aggressive humans? In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, (pp. 553–561). International Foundation for Autonomous Agents and Multiagent Systems.

Bosse, T., Gerritsen, C., de Man, J., & Treur, J. (2013). Effects of virtual training on emotional response. In K. Imamura, S. Usui, T. Shirao, T. Kasamatsu, L. Schwabe, & N. Zhong (Eds.) *Brain and Health Informatics*, vol. 8211 of *Lecture Notes in Computer Science*, (pp. 21–32). Springer International Publishing.

Bouchard, S., Guitard, T., Bernier, F., & Robillard, G. (2011). Virtual reality and the training of military personnel to cope with acute stressors. In *Advanced Computational Intelligence Paradigms in Healthcare 6. Virtual Reality in Psychotherapy, Rehabilitation, and Assessment*, (pp. 109–128). Springer.

Brouwer, A.-M., Neerincx, M. A., Kallen, V., van der Leer, L., & ten Brinke, M. (2011). Eeg alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress. *Journal of Cybertherapy & Rehabilitation*, *4*(1), 21–34.

Courtney, C. G., Dawson, M. E., Schell, A. M., Iyer, A., & Parsons, T. D. (2010). Better than the real thing: Eliciting fear with moving and static computer-generated stimuli. *International Journal of Psychophysiology*, *78*(2), 107–114.

Craig, K. D., & Lowery, H. J. (1969). Heart-rate components of conditioned vicarious autonomic responses. *Journal of Personality and Social Psychology*, *11*(4), 381–387.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210.

Epstein, S., & Roupenian, A. (1970). Heart rate and skin conductance during experimentally induced anxiety: The effect of uncertainty about receiving a noxious stimulus. *Journal of Personality and Social Psychology*, *16*(1), 20–28.

Fowles, D. (1982). Heart rate as an index of anxiety: Failure of a hypothesis. *Perspectives in cardiovascular psychophysiology*, (pp. 93–126).

Goldstein, M. J., Jones, R. B., Clemens, T. L., Flagg, G. W., & Alexander, F. G. (1965). Coping style as a factor in psychophysiological response to a tension-arousing film. *Journal of Personality and Social Psychology*, *1*(4), 290–302.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(03), 281–291.

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. G., Neerincx, M. A., & Brinkman, W.-P. (2014). Controlling social stress in virtual reality environments. *PloS one*, *9*(3), e92804.

Hassett, J. (1978). *A primer of psychophysiology*. WH Freeman.

Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience letters*, *328*(2), 145–149.

Lazarus, R. S., Speisman, J. C., & Mordkoff, A. M. (1963). The relationship between autonomic indicators of psychological stress: Heart rate and skin conductance. *Psychosomatic Medicine*, *25*(1), 19–30.

Lin, H.-P., Lin, H.-Y., Lin, W.-L., & Huang, A. C.-W. (2011). Effects of stress, depression, and their interaction on heart rate, skin conductance, finger temperature, and respiratory rate: sympathetic-parasympathetic hypothesis of stress and depression. *Journal of clinical psychology*, *67*(10), 1080–1091.

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.) *Handbook of affective sciences*, (pp. 619–642). Oxford University Press.

Markovsky, B. (1988). Injustice and arousal. *Social Justice Research*, *2*(3), 223–233.

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform.* Muska & Lipman/Premier-Trade.

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and social psychology bulletin*, *24*(7), 769–783.

Nogueira, P. A., Rodrigues, R., Oliveira, E. C., & Nacke, L. E. (2013). Guided emotional state regulation: Understanding and shaping players' affective experiences in digital games. In *Proceedings of the Ninth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, (pp. 51–57). AAAI.

Ohman, A. (2000). Fear and anxiety: Evolutionary, cognitive, and clinical perspectives. *Handbook of emotions*, *2*, 573–593.

Ozel, F. (2001). Time pressure and stress as a factor during emergency egress. *Safety Science*, *38*(2), 95–107.

Speisman, J. C., Lazarus, R. S., Mordkoff, A., & Davison, L. (1964). Experimental reduction of stress based on ego-defense theory. *The Journal of Abnormal and Social Psychology*, *68*(4), 367–380.

Steuer, J. (1995). Defining virtual reality: Dimensions determining telepresence. In F. Biocca, & M. Levy (Eds.) *Communication in the Age of Virtual Reality*, (pp. 33–56). Lawrence Erlbaum Associates.