Meta-analysis of digital game and study characteristics eliciting physiological stress responses

BENNY VAN DER VIJGH,[a,b] ROBBERT-JAN BEUN,[a] MAARTEN VAN ROOD,[a,b] AND PETER WERKHOVEN[a]

[a]Buys Ballot Laboratory, Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
[b]Stratenum, Department of Neurology & Neurosurgery, University Medical Center Utrecht, Utrecht, The Netherlands

Abstract

Digital games have been used as stressors in a range of disciplines for decades. Nonetheless, the underlying characteristics of these stressors and the study in which the stressor was applied are generally not recognized for their moderating effect on the measured physiological stress responses. We have therefore conducted a meta-analysis that analyzes the effects of characteristics of digital game stressors and study design on heart rate, systolic and diastolic blood pressure, in studies carried out from 1976 to 2012. In order to assess the differing quality between study designs, a new scale is developed and presented, coined *reliability of effect size*. The results show specific and consistent moderating functions of both game and study characteristics, on average accounting for around 43%, and in certain cases up to 57% of the variance found in physiological stress responses. Possible cognitive and physiological processes underlying these moderating functions are discussed, and a new model integrating these processes with the moderating functions is presented. These findings indicate that a digital game stressor does not act as a stressor by virtue of being a game, but rather derives its stressor function from its characteristics and the methodology in which it is used. This finding, together with the size of the associated moderations, indicates the need for a standardization of digital game stressors.

Descriptors: Physiological stress response, Digital game, Stressor, Meta-analysis, Game characteristics, Study characteristics

Digital games have been widely used as psychological stressors in the past decades and have been found to induce both physiological and psychological stress responses, such as altered heart rate and blood pressures increased cortisol level, altered aggressiveness, and prosocial behavior (see, among others, Barlett, Anderson, & Swing, 2009; Barlett, Harris, & Baldassaro, 2007; Dickerson & Kemeny, 2004; Miller & Ditto, 1988; Saab et al., 1991). They act as a psychological stressor by virtue of the challenges imposed on individuals: digital games present complex tasks that require intensive mental processing, are able to produce aversive stimuli, or create frustration from the inability to reach a desired goal in the game. The most important advantages of using a digital game as a stressor over other, more conventional, psychological stressors such as mental arithmetic, public speaking, or picture-rating tasks are the flexibility of, and the precise control over, the stressor. Digital games are flexible because different forms of psychological stressors, such as workload, social-evaluative threat, and frustration, can all be presented within a single setting. Additionally, due to the digital nature of the stressor, digital games allow for precise control of these stimuli, because the output can be determined in detail and manipulated in real time.

The complexity underlying this flexibility and precise control requires a one-time thorough analysis, however, in order to correctly interpret the observed physiological stress responses. For example, one of the most assessed physiological responses in research utilizing digital games as a stressor is the cardiovascular response, such as the heart rate (HR), systolic (SBP), or diastolic blood pressure (DBP) response. In a variety of studies, these cardiovascular measures are determined before, during, and/or after playing a digital game. The general response is an increase in all three cardiovascular measures, yet the magnitude of these responses varies greatly between studies (see, e.g., Lawler, Wilcox, & Anderson, 1995; Sita & Miller, 1996—the latter reports effect sizes for HR over four times higher than the former). In order to understand these differences and hereby gain insight in the underlying stress response systems, the digital game acting as a stressor needs to be characterized and analyzed.

Although there exist numerous reviews and meta-analyses on the effects of digital games, to our knowledge no meta-analysis has

yet made a thorough examination of the relation between characteristics of digital game stressors and physiological responses to these stressors. Many of the reviews and meta-analyses on the effects of digital games focus on the effects of digital games on psychological and behavioral constructs, such as aggression, violence, and prosocial behavior (see, among others, Anderson & Bushman, 2001; Barlett, Anderson, & Swing, 2009; Huesmann & Bushman, 2006; Sherry, 2001). Two meta-analyses do examine the effect of digital games on physiological responses and find that (violent) digital games have a positive correlation with physiological arousal ($r =$ .21, Barlett, Anderson, & Swing, 2009; and $r = .26$, Huesmann & Bushman, 2006). Although these studies note that different game characteristics show similarity with types of stressors known to have physiological effects outside the context of a digital game, for example, aversive stimuli from standardized stimuli sets such as the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) and the International Affective Digitized Sounds (IADS; Bradley & Lang, 1999), neither of these studies explicitly coded specific game characteristics. As a result, it is currently unclear which characteristics of digital game stressors are related to the physiological stress responses elicited by these stressors. This makes it difficult, if not impossible, to interpret and compare the results from these stressors and to gain insight in the underlying stress response systems.

In addition, study characteristics (not related to the game itself) may also have significant effects on the stress response. An example of such a study characteristic is social-evaluative threat, induced by the presence of either cameras or the experimenter during the experiment, being a significant moderator of the cortisol stress response (Dickerson & Kemeny, 2004).

Stress and Digital Games

To present a single and consistent definition that provides sufficient generality to capture the different approaches to stress found in the stress literature, we adhere to a definition based on the work by Newport and Nemeroff, 2002, who considered stress as any challenge to the homeostasis of an individual that requires an adaptive response of that individual. We utilize this definition in the following form: Stress is the state resulting from the ensemble of responses that are aimed at (facilitating) restoration and/or maintenance of (psychological) homeostasis to internal or external stimuli that present (perceived) challenges to (psychological) homeostasis.

To further refine our definition, we will refer to the stimulus presenting the challenge to (psychological) homeostasis as the *stressor*, and the response to this challenge as the *stress response*. Consistent with the interactional approach on stress (Jones & Bright, 2001), we identify variables influencing the relation between the stressor and stress response as intervening variables. Intervening variables include categories such as age, sex, and personality type as well as, for example, categories including environmental variables such as temperature.

Using the definition by Adams (2010), we define a game as a type of play activity, conducted in the context of a pretended reality, in which one or more players tries to achieve at least one nonarbitrary, nontrivial goal by acting in accordance with rules. Key elements in games are players, (inter)action, environment, goals, and rules. Within digital games, key elements are represented by computer technology. A computer may simulate (parts of) the game environment, the interaction, or one or more players. In single-player games, only one human player is involved in the

game at any one time; in multiplayer games, a computer often functions as a complex intermediary between human players.

A player interacts with a digital game by means of its user interface. A user interface receives user input through one or more input modalities, for example, through human forces exerted on a keyboard and mouse. In response, the user interface presents the output, perceptual stimuli, to the user, such as images, sounds, music, and vibrations.

Two key concepts in user interfaces are the interaction model and its perspective (Adams, 2010). The interaction model determines how human players impose their desired actions via the input devices on the game. The perspective refers to the camera view of the game, for instance, first person, third person, or top-down perspective. Interaction models can be classified in various types, such as avatar based or omnipresent. In an avatar-based model, a player acts upon the world through a single character, the avatar. In the omnipresent model, a player can act upon different parts of the game world at the same time due to the omnipresence of the player, often visualized through an aerial perspective. In this meta-analysis, we will refer to game characteristics as any (or part) of the key elements or the game as a whole, such as the game type, the presence of game music, or the perspective used in the digital game.

Based on the divergent physiological responses to digital game stressors (e.g., Lawler et al., 1995, vs. Sita & Miller, 1996), the effects of study characteristics on the stress response(Dickerson & Kemeny, 2004), and the notion of game characteristics showing similarities with stressors known to have physiological effects outside the context of a digital game (Barlett, Anderson, & Swing, 2009; Huesmann & Bushman, 2006), we hypothesize that the characteristics of both game and studies are significant moderators of the physiological stress responses to these digital game stressors. Therefore, in this paper, we will conduct a meta-analysis that aims to analyze whether characteristics of digital games and studies are related to the physiological stress response elicited by these stressors, and examine the nature of these relations.

Method

We conducted a computerized search in four relevant digital databases, PubMed, Scopus, PsycInfo, and the IEEE Xplore Digital Library, using the search phrases *game and physiolog** and *stress* and game*. Titles, abstracts, and keywords were queried for these search phrases, in which the word *and* represented logical conjunction and the * character (wild card) represented any character to be filled in at this position, any number of times, allowing for the construction of words such as *physiological*, *physiology*, or *stressor*. PubMed and Scopus covered publication dates from 1966 onwards, PsycInfo and IEEE Xplore covered, depending on the specific fields, the 20th century and onwards, all as indexed by March 2012, the month in which the queries were executed.

This search generated 5,448 articles (728 articles from PubMed, 3,866 from Scopus, 636 from PsycInfo, and 208 from the IEEE Xplore Digital Library). The abstract of each article was examined, and all articles that matched the six inclusion criteria below were obtained to match the complete article against these criteria. Furthermore, to identify additional articles not detected using the computerized search, an iterative exploration of the reference lists of the included articles was performed. Although the precise creation date of the first digital game is open for debate, with regard to digital games that are used for stress research it can safely be assumed that the creation of the first such digital game did not take place

before 1966. Because of this, no additional journals were searched by hand to supplement this computerized search, since it fully covered the range of possible publication years of research utilizing digital games as stressors.

The used inclusion criteria are given below.

1. Inclusion was limited to studies that utilized a single-player digital game as stressor, with the exclusion of exergames or games utilizing bio- or neurofeedback.

   Multiplayer digital games are excluded because different stress responses in HR and electrodermal activity have been observed between participants playing either against the computer or human counterpart (Lim & Reeves, 2010). Comparable results have been found by Ravaja (2009) and van Egeren (1983), as well as outside the context of digital games, where Verona, Sadeh, and Curtin (2009) observed heightened responses to an air puff in the face, when this was administered by a human instead of a machine. Digital games utilizing bio- or neurofeedback are excluded here, since the feedback loop between the participant and game utilized in these games potentially influences the observed stress responses (see Goodie & Larkin, 2001). Exergames, that is, games that are also a form of exercise during execution, such as games controlled through Microsoft's Kinect, are excluded since the exercise itself acts as a form of physical stressor, influencing the observed stress response.

2. The study needed to measure physiological stress responses during and/or after stressor presentation, as well as before stressor onset (so-called baseline measurements). This had to be performed in a within-subject design in order to be able to calculate the effect size.

3. Between the baseline measurement and the stress response measurement, participants should be exposed to only a single digital game stressor (as opposed to multiple digital game stressors) in order to uniquely relate the stress response to the characteristics of this specific digital game.

   There were three exceptions made to this criterion: Miller & Ditto (1988), Miller & Ditto (1989), and Miller, Friese, & Sita (1995). All three studies used the same procedure in which the same four digital games were presented in succession between baseline and response measurements. Because the games did not differ in any of the coded game characteristics, the studies were included.

4. The study must include healthy, adult participants. Studies need to report participants to be in good health, assessed either through self-report or standardized relevant medical checks. Of studies including participants with a reported physical or psychological diagnosis (e.g., posttraumatic stress disorder, depression, hypertension, or diabetes), only healthy control groups were included. Experimental groups in which the cognitive abilities or physiological response systems of participants were potentially altered (e.g., through hypnosis or through chemical agents such as ethanol or sodium intake prior to or during stressor presentation) were excluded. With regard to the age criterion, studies were included if the reported age range of the participants did not include ages under 18. When an age range was not given, inclusion was based on the description of the participants, such as "graduates" or "managers." Descriptions referring to demographic groups generally consisting of an adult population were admitted, including both graduates and undergraduates. When neither the age range nor description of the participants was given, inclusion was based on the reported mean age and standard deviation, which we approached in a conservative manner:

when the mean age minus the standard deviation was below 18, the study was excluded. When none of these three statistics was given, the study was also excluded.

5. In order to exclude pilot studies and studies with very low statistical power, the experimental group that is to be included must contain at least 10 participants. Pilot studies have, by their definition, a different nature than regular studies (i.e., to test certain aspects of a study design, such as its methodology or stimulus). We aimed to include only studies that used a design that was considered by the authors to be a regular study, and not pilot studies in which the design had not yet fully matured.

6. The study must be reported in a journal article, published in an English language, peer-reviewed journal.

The above-described method resulted in a final study sample of 54 articles containing 66 studies, spanning the years 1976 until 2012 and containing 3,863 participants.

Coded Characteristics

In order to determine which game and study characteristics will be coded in the acquired study sample, we identified characteristics as potential predictors of physiological stress responses based on findings in the literature outside of the study sample. We initially identified 18 potentially relevant game characteristics and seven relevant study characteristics (these characteristics are given in online supporting information Table A1). Of these characteristics, five game and one study characteristic were excluded because these variables were either not measured and/or reported in a considerable portion of the study sample, or were reported in an inconsistent manner across studies, making it impossible to quantify these characteristics across studies. The remaining 19 characteristics were used for coding of the 66 studies in the study sample. (A table containing all of these studies coded on all 19 characteristics is not included due to space limitations, but is available upon request from the corresponding author.)

After coding, the values found for nine of the characteristics (eight game and one study characteristic: the platform on which the game is executed, the absence or presence of external influence on stressor, the used output modalities, amount of control over reaching set goal, absence or presence of time pressure, absence or presence of input/output frustration, color scheme used, sound level, and the measurement method used) did not exhibit sufficient variance across studies. Based on the rule of thumb that at least 10 studies are needed to include a characteristic in later analyses (A. Field, 2009), we set the cutoff for variance at minimal 10 studies exhibiting unique values for continuous variables and minimal 10 studies exhibiting values of each class for discrete variables. Therefore, to avoid deriving spurious relations, these characteristics were excluded from the analysis. The remaining five game and five study characteristics are used in the analysis. A detailed treatment of these characteristics is given below.

Game Characteristics

Game type. This characteristic refers to the type of digital game used as stressor. Examples of different game types include action, adventure, strategy and management, role-playing games, simulation, or board and card games (Ritterfeld, 2009). Because all digital games included in this meta-analysis could be classified as action games, we used a further subdivision of this specific game type as defined by Adams (2010), coding each game as either a 3D

shooter, 2D shooter (i.e., a game in which the main goal can be achieved by shooting, either in a 3D or 2D environment), fast puzzle, racing/sports, platform (i.e., a game in which the environment is depicted in a side view and the avatar can be moved from one side of the screen to the other, while the screen scrolls in the direction of the movement), or a fighting game.

Input modalities. The different modalities used to provide input to the game were coded. Because all games in the study sample used the tactile modality as input modality, we made a further subdivision based on the specific input devices used in our study sample. Three subtypes were coded: tactile feedback controllers (i.e., controllers that can vibrate during execution of the game), joysticks or paddles (mostly used by older consoles such as the Atari 2600), and the keyboard and mouse combination.

Aversive stimuli. Aversive stimuli can elicit physiological stress responses. The presentation of aversive stimuli such as in a picture-rating task (Stegeren, Wolf, & Kindt, 2008), passive viewing of aversive pictures (Sokhadze, 2007), or film viewing containing aversive stimuli (Miller, 1993) has consistently been shown to induce physiological stress responses such as changes in HR, electrodermal activity, and frontal EEG activity. Within the context of a game, aversive stimuli can include, for example, the presence of violence, blood and gore (Hebert, Beland, Dionne-Fournelle, Crete, & Lupien, 2005), and torture (Tafalla, 2007). These stimuli within a digital game have been shown to induce physiological stress responses (Carnagey, Anderson, & Bushman, 2007).

Our coding of aversive stimuli is based on the stimuli observed in the study sample and ordered on the basis of the intensity of the stimulus, using recommendations made by two advisory boards, the Pan European Game Information organization (PEGI, 2012) and the Entertainment Software Rating Board (ESRB, 2012). These boards test digital games and provide suggestions on which age category is suited for the specific game. Six categories were distinguished: (1) supernatural entities and related imagery (e.g., images of devils or gods), (2) disturbing sounds without sight of source, (3) violence (e.g., beating of other human beings), (4) blood and gore (e.g., intestines being visible or horrid mutations), (5) killing, and (6) torture (e.g., physical abuse of a restrained individual). Based on ratings given by the advisory boards for these or comparable categories, we assigned Categories 1, 2, and 3 one point regarding intensity, Categories 4 and 5 two points, and Category 6 three points. Furthermore, based on indications that aversive stimuli can have a cumulative effect (cf. Barlett, Harris, & Bruey, 2008), we coded each of the six different aversive categories that were present in the game and summed the corresponding intensity values to calculate an overall aversiveness score for the digital game used in that study. In this manner, the overall score lies between 0 and 10.

Realism. Several studies have found the amount of realism in digital games to be related to physiological stress responses to these games. For example, Ivory and Kalyanaraman (2007) observed that more technologically advanced, although otherwise comparable, digital games elicited higher electrodermal stress responses. Similarly, Barlett and Rodeheffer (2009) showed that more realistic digital games significantly heighten the HR stress response compared to less realistic games. Studies examining this game characteristic often employ different definitions of realism, such as the probability of seeing an event presented in the digital game in real life (Barlett & Rodeheffer, 2009) or by using the year of release of the

game (Ivory & Kalyanaraman, 2007), as digital games become more and more realistic with respect to stimuli presentation due to the continuously developing graphic and game design abilities.

To allow comparison of realism across studies in our study sample while aiming to capture the multidimensional nature of this construct, we coded three relevant elements that are objectively quantifiable. The first element concerns the presence of an avatar in the game and, if present, whether the avatar has a human form. This has been shown to be a significant contributor to the feeling of a participant that the representations offered by the digital game are real (cf. Lee, 2004; Lombard & Ditton,1997). Slater and Usoh (1993) showed that participants reported significantly higher presence when controlling an avatar of their own body compared to controlling a nonhuman cursor. The second element concerns the perspective employed in the game. Participants report greater presence when the perspective is in first person, that is, through the eyes of the avatar, compared to a third person perspective (Dahlquist, Herbert, Weiss, & Jimeno, 2010). The third element is the year of release of the digital game, as used by Ivory and Kalyanaraman (2007).

We assigned each element a value between 0 and 2, as this corresponded with the number of classes in two of the elements, and summed these values. We had no theoretical grounds to give any of the three categories more weight in the summed realism, so the three categories had the same weights in this summation. Presence of an avatar was scored as 0 in cases when no avatar was present in the game, 1 if there was a nonhuman avatar, and 2 if a human avatar was provided in the game. Perspective was scored as 0 when a side view or top-down perspective was employed, 1 point was scored when a third person perspective was employed, and 2 points if a first person perspective was used. To normalize year of release to a value between 0 and 2 as well, the release year of a digital game was normalized over the 34 years spanning the earliest release date of a digital game found in our study sample, 1976, and the latest release date, 2010. This was done by subtracting the release date of a given digital game by 1976, and subsequently dividing this by 17 (half of 34) to achieve a normalized value between 0 and 2.The summation of the values of the three different elements results in an overall scale between 0 and 6, with higher values indicating higher realism.

Game music. A recent review by Sokhadze (2007) makes it clear that music has the potential to elicit physiological (stress) responses, although there are inconsistent results to be found regarding the direction of the elicited responses. Examples include work by Nyklicek, Thayer, and van Doornen (1997), who observed significant changes in both cardiovascular and respiratory variables in response to different fragments of music and white noise, and McFarland (1985), who found that music with different valence and arousal ratings has different effects on skin temperature. As it proved difficult to objectively determine the type of music presented, we solely coded whether music was present in the game.

Study Characteristics

Additional external stressors. Additional external stressors are stressors outside the digital game that are presented to the participant within a relevant time window around the baseline and stress response measurements. The size of this window differs per physiological stress response. For example, in the case of heart rate variability, the relevant time window is relatively small since heart rate variability reacts quickly to presented stressors. Virtually all

external stressors coded in the study sample were presented during the stressor presentation, that is, during the playing of the game, which is a relevant time window for every stress response measure included in this meta-analysis.

These stressors can influence the measured stress response and may obscure the relation between game or study characteristics and their impact on physiological activity. We identified four external stressors in our study sample: social-evaluative threat, inclusion of stressor in task battery, the use of incentives or shocks, and the use of loud noise.

Social-evaluative threat is the possibility that a participant could be negatively judged by others (Dickerson & Kemeny, 2004). A study is coded as containing social-evaluative threat if the experimenter stayed in the room during execution of the game by the participant, when there was a camera present during the experiment, or if the experimenter informed the participant that he or she is performing worse than his or her peers.

The second additional external stressor is the inclusion of the digital game stressor in a battery of tasks, that is, a set of tasks or stressors presented to the participant in a series, in contrast to the presentation of one stressor. This procedure is found in several studies, mostly utilizing additional tasks such as reaction time or speech tasks (e.g., Ironson et al., 1989; Saab et al., 1991; Sita & Miller, 1996). This inclusion may alter the obtained baseline of the digital game stressor in cases where the other stressors are presented before obtaining the baseline, or change unmeasured internal endocrine or other physiological levels, which, in turn, potentially alter the stress response through carryover effects. When the inclusion in a task battery is controlled for in the study by treating task order as a covariate in the analysis and yielding nonsignificant results, the inclusion in a task battery is still coded as an additional stressor in the analysis. This is because this return to baseline does not guarantee the normalization of other internal endocrine or physiological levels potentially altering the stress response.

The third external stressor coded in our study sample consists of giving the participants either an incentive or a shock, contingent on performance. These external stressors have been shown to influence the physiological stress response (see, among others, Lovallo, 2005; Miller, 1993; Miller & Ditto, 1988; Perkins, Leonard, Jennings, & Stiller, 1986).

The final external stressor identified is the presentation of a loud noise. Several studies have implicated high sound levels (mostly 75 dB and above) with physiological (and psychological) stress responses (Selander, et al., 2009; Smith, Whitney, Thomas, Perry, & Brockman, 1997). We coded the presentation of loud noise as an external stressor when it was reported that auditory stimuli exceeded 75dB or when the presence of "loud noise" (or comparable wording) was reported.

To account for the cumulative effects of external stressors (Dickerson & Kemeny, 2004), we took the sum of all additional external stressors coded for each study. Because we had no theoretical grounds to give any of the different external stressors more weight in this sum, the categories had the same weights in this summation, each providing one point.

Intervening Variables

A plethora of variables have been shown to moderate stressor-stress response relations, referred to as intervening variables in this analysis. However, the frequency of reporting these variables is low for the studies included in this analysis. Therefore, we coded only the most widely reported intervening variables. These are age,

sex, sociodemographic group, race, psychological and medical disorders (including disorders of parents), medication, intake of caffeine or nicotine, education, whether the participant is a smoker or not, the position in which the game was played, whether the experimenter had the same sex as the participant, and digital gaming frequency (i.e., the frequency with which participants use digital games in daily life). Of these variables, only age and sex were reported sufficiently, as well as varying substantially across the study sample, and are therefore used in the analysis.

Age and sex have been found to be significant moderators of physiological stress responses in several studies. Matthews and Stoney (1988) observed that, using acute psychological stressors (serial subtraction and mirror-image tracing), adults had increased blood pressure responses compared to children, whereas HR responses showed a negative correlation with age. Furthermore, several studies found that in general males show increased blood pressure responses. Examples are increased responses to a serial subtraction task, speech task, or self-evaluation as found by Stoney et al., (1988), or to mirror-image tracing as observed by Matthews & Stoney (1988), or to a shock-avoidance reaction time task, as described by Hastrup and Light (1984). Stoney et al. (1988) also found that females showed increased HR responses to a speech task.

Therefore, in cases where studies reported stress response measures for males and females separately, the different sex groups are admitted as separate studies to gain insight in the effect of sex on the stress response. For age, we coded the mean age of the study. If this was not given, we used age range, of which we used the value in the middle of the range. If no age was given, we used sociodemographic group information and coded the age as the mean that the specific sociodemographic group had in our sample. For example, the mean age of students in our sample was 21.3, so the age of any studies that used students as participants and not reported a mean age or age range was coded as 21.3.

Effect Size Reliability

Several factors influence the reliability of the effect size calculated for each study. These factors include the differing quality of the methodologies employed in the study, the reliability of the different ways in which data were extracted from the studies, and the manner in which effect sizes were calculated. Therefore, we created a new scale, encompassing 10 of such factors. These factors take into account general research methodological demands, such as drafted by Cook and Campbell (1979), and domain-specific factors that are relevant in the field of physiological stress research. Every factor, if unsatisfied, subtracts one point from a starting score of 12 points, with a higher score indicating a higher reliability of the effect size calculated for the respective study. These factors are given below.

1. Substance refrainment. When participants were not asked to refrain from caffeine or nicotine at least 4 hr prior to the experiment, subtract one point. These substances act on the same physiological systems underlying many of the measured stress responses, which can obscure the stressor-stress response relation.
2. Measurement location. If measurements are made on the dominant arm, subtract one point. If no indication of dominance is given, right is assumed to be dominant. Measurements on the dominant arm may potentially cause artifacts in the measured signals due to movement of this arm during the experiment.
3. Stressor onset timing. If there is a time interval between the end of baseline measurements and stressor onset,

subtract one point. In this interval, physiological changes can occur rendering the baseline measurement unrepresentative of the physiological state of the participant at stressor onset.

4. Baseline conditions. When participants were required to perform actions during baseline measurements such as filling out questionnaires, subtract one point. These actions potentially interfere with the baseline measure.

5. Baseline duration. If the duration of the baseline measurement is under 3 min, subtract one point. This is a minimum time required for most measured physiological responses to reach a baseline value, with the exception of hormonal responses.

6. Response measurement timing. If the stress response is measured after instead of during stressor presentation, with the exception of endocrine measures, subtract one point. Measurements obtained after the stressor presentation are less representative of the stress response compared to measurements obtained during the response.

7. Number of measurements. When only one measurement was made during either or both baseline and stress response measures, subtract one point. One measurement provides less reliability of acquiring a value representative of the actual state of the measured physiological signal, compared to multiple measurements.

8. Presence of preceding tasks. If tasks were presented prior to stressor onset, subtract one point. Preceding tasks can alter the measured stressor-stress response relation.

9. Calculation values. When calculations are based on nonexact values of p (e.g., reporting of significance in the form $p < .05$), subtract one point. When a study reports "(no) significance," subtract one additional point.

10. Calculation methods. In case any nonstandard calculation methods have been employed (i.e., methods other than using the exact measurement values or exact p, t, or F statistics from the respective statistical tests, such as deriving the values from a plot), subtract one point. 10.

Stressor Duration

As was shown by Barlett and Rodeheffer (2009), physiological stress responses can show different patterns over different time periods. Therefore, we coded the duration the stressor was presented to the participants in minutes, (i.e., the time the participants played the stressor game).

Number of Participants

We coded the number of participants included in the study that was used in the calculation of the reported stress response values.

Coding of characteristics. After training, using a randomly selected set of studies included in the analysis, two researchers independently coded the studies according to an established protocol. This protocol was drafted specifically for this analysis and maintained throughout the coding process. Any differences discovered during the coding process were discussed, and the protocol was updated to reflect the decided coding strategy. The agreement rate between the studies coded by the coders was 93%, determined on a randomly selected set of 10% of the study sample.

In studies presenting digital game stressors that differed in game characteristics between subject groups, we coded the differ-

ent groups as separate studies. This allows us to gain insight in the effect of these characteristics on the stress response. For example, in the study by Barlett and Rodeheffer (2009), the participants in the different experimental groups played dissimilar digital games, differing in several relevant game characteristics such as the realism of the game and the presence of aversive stimuli in the game. These different groups are coded as separate studies, allowing insight in the effect of, in this case, realism and aversive stimuli on the physiological stress response. When the between-subject factor was not based on a game characteristic, such as in the case where groups are based on the presence of parental hypertension, these groups were collapsed, and the data of the complete study population was used in order to optimally use the amount of data available in the study sample.

In the case of a study using a within-subjects design with the levels of the within-subject factor being manipulations of a coded characteristic, the different values of this factor are not coded as separate studies since this would violate the independence between effect sizes assumed by our analysis methods. This occurred in one study (Carroll, Turner, Lee, & Stephenson 1984), where participants were presented with the stressor protocol in two separate sessions, one week apart. In this case, we only included the first session as a study, to control for potential learning and habituation effects present in measurements of the second session.

In case a relevant characteristic was not explicitly reported, an effort was undertaken to derive the characteristic indirectly from other information given in the study. For certain characteristics, when no explicit value was provided, the standard value of this characteristic, (i.e., the value without any alteration) was assumed. For example, for games offering differing levels of aversiveness (e.g., the possibility to either display blood or not), we assumed that the game was played without any alteration, being the standard setting provided by the game. All derived or assumed standard values are given in italics in Table 1, containing the study sample used for analysis. When it was not possible to derive or use the standard value of a characteristic, the corresponding author of the study was contacted to retrieve the value.

Effect Size Calculation

To calculate the effect size for all the physiological stress responses encountered in the study sample, we examined the standardized mean difference effect between the baseline and stress response values, using the unbiased version of Hedges's g, denoted as $g*$ (Hedges, 1981). The formulas for calculating g and $g*$ are given by:

$$g = \frac{\mu_{\text{stress response}} - \mu_{\text{baseline}}}{\sigma_{\text{baseline}}}$$

$$g* = \left(1 - \frac{3}{4(2*n) - 9}\right) * g$$

Here, $\mu_{\text{stress response}}$ and μ_{baseline} refer to the mean value of the measured physiological signal during the stress response measure and baseline measure, respectively. Furthermore, σ_{baseline} refers to the standard deviation of the baseline measurement values, and n stands for the number of participants in the respective study. In this manner, the sign of the resulting g or $g*$ will be positive when the physiological stress response is higher compared to the corresponding baseline measure, and the value indicates the change from the baseline, expressed in standard deviations. Because g has a bias (i.e., with few participants the calculated effect size is

Table 1. *Study Sample Used for Analysis*

| | Game characteristics | | | | Study characteristics | | | | | | Effect size $g*$ | | |
Study	Game type	Input modalities	Aversive stimuli	Realism	Game music	Additional external stressors	Gender	Age	Reliability of effect sizes	Stressor duration	N	HR	SBP	DBP
Arriaga, 2006	3	*3*	0	2.03	0	0	−1	23	11	2	87	0.28		
Baldaro, 2004.1	1	3	4	5.48	*0*	0	1	25	11	20	11	1.07	0.77	0.17
Baldaro, 2004.2	3	3	0	2.23	*1*	0	1	25	11	20	11	0.22	−0.13	−0.27
Barlett, 2009a.1	1	*1*	3	4.74	*1*	0	−1	22	7	15	26	0.38		
Barlett, 2009a.2	1	*1*	2	3.78	*1*	0	−1	22	7	15	25	−0.15		
Barlett, 2009a.3	4	*1*	0	4.68	*1*	0	−1	22	7	15	23	−0.59		
Barlett, 2008.1	6	*1*	6	4.68	1	0	−1	19	7	15	*19*	0.31		
Barlett, 2008.2	6	*1*	5	4.68	1	0	−1	19	7	15	56	0.15		
Barlett, 2009b.1	6	1	6	4.68	1	1	−1	19	8	15	47	0.68		
Barlett, 2009b.2	4	1	0	4.68	1	1	−1	19	8	15	44	0.09		
Dembroski, 1978	3	*2*	0	*1.06*	0	1	1	*20*	10	3	50	0.55	0.91	0.79
Dembroski, 1985	3	2	0	2	*0*	2	0	21	11	12	10	−0.73	−0.40	−0.09
Ditto, 1989	4	*2*	0	2.32	0	1	1	24	10	15	10	0.40	0.95	0.95
Goodie, 2001	4	2	0	3.52	0	1	1	*21*	11	4	13	1.09	0.66	0.79
Grossman, 1987	3	2	0	1.71	0	2	−1	*21*	9	3	19	1.55		
Ironson, 1989	3	2	0	1.39	*0*	0	−1	34	10	10	119		0.90	0.85
Lawler, 1995.1	2	2	0	4.52	*0*	1	0	*21*	9	4	55	0.47	1.19	0.58
Lawler, 1995.2	2	2	0	4.52	*0*	1	1	*21*	9	4	64	0.65	1.21	0.88
Markovitz, 1998.1	3	2	0	1	*0*	0	0	27	8	3	906		1.03	1.02
Markovitz, 1998.2	3	2	0	1	*0*	0	1	27	8	3	899		1.15	1.0
McKinney, 1985	3	2	0	1	*0*	1	1	45	10	5	59	1.06	1.40	1.05
Miller, 1988	4	2	0	2.32	0	1	1	*21*	10	15	14	1.50	3.67	3.17
Miller, 1989	4	2	0	2.32	0	1	1	*21*	10	15	32	1.48	1.39	1.91
Miller, 1991	4	2	0	2.32	0	1	1	23	10	15	24	1.43	1.12	2.45
Miller, 1992	4	2	0	2.32	0	1	1	23	11	15	48	1.83	0	0
Miller, 1993	4	2	0	2.45	*0*	1	1	*21*	10	3	40	0.92	1.34	1.71
Miller, 1994	4	2	0	2.45	0	1	1	22	10	3	24	0.89		
Miller, 1995	*4*	2	0	*2.32*	0	1	1	24	11	15	36	1.43	1.02	0.65
Perkins, 1994	4	2	0	2.45	0	1	1	22	10	10	12	1.29	2.07	1.09
Pollak, 1994	2	2	0	1.06	*0*	0	1	28	9	5	26	1.44		
Saab, 1991	3	2	0	2.39	*0*	0	1	35	9	10	23	8.54		2.76
Sims, 1990	*3*	2	0	2.65	0	1	1	21	10	4	25	1.60	1.23	0.79
Sita, 1996	4	*2*	0	2.45	0	2	0	27	11	3	30	1.96	1.46	1.75
Staude-Muller, 2008.1	1	*3*	4	5.67	1	0	1	22	10	20	20	0.01		
Staude-Muller, 2008.2	1	*3*	0	5.67	1	0	1	22	10	20	21	−0.28		
Steptoe, 1984	−1	2	0	1.52	0	0	1	43	9	4	12	0.21	0.95	0.82
Tafalla, 2007.1	1	3	6	5.1	1	0	−1	21	10	30	35	1.02	0.39	1.39
Tafalla, 2007.2	1	3	6	5.1	0	0	−1	21	10	30	38	0.63	0.04	0.79
Tischenkel, 1989.1	3	2	0	2.39	0	0	1	*35*	10	10	22	0.44	0.44	0.44
Tischenkel, 1989.2	3	2	0	2.39	0	0	0	*35*	10	10	17	0.44	0.44	0.44
Turner, 1983	2	3	0	1.06	1	0	1	*23*	11	10	24	0.87		
Ward, 1986	3	*2*	0	1	*0*	0	1	50	8	−1	56	0.29	0.10	0.02
Zachariae, 2000	3	3	0	*1*	0	2	−1	28	10	20	26			

Note. Studies are referenced using the name of the first author. All derived or assumed values are given in italics. DBP = diastolic blood pressure; heart rate = HR; systolic blood pressure = SBP. Coding for game type: 1 = 3D shooter; 2 = 2D shooter; 3 = fast puzzle game type; 4 = racing/sports game type; 5 = platform; 6 = fighting game. Coding for input modalities: 1 = tactile feedback controller; 2 = paddle/joystick; 3 = keyboard mouse. Coding of game music: no music = 0; music = 1. Coding for sex: male = 1; female = 0. N = number of participants included in the analysis of the respective study. For all characteristics, the coding of -1 means unknown data.

overestimated), we calculate $g*$ in which this bias is corrected (DeCoster, 2009). This effect size $g*$ is used for all analyses. In cases where the mean of the baseline and response measurements and accompanying standard deviation of the baseline were not available, we used standard meta-analytic procedures, such as t or F statistics or exact values of p for specific statistical tests to calculate $g*$ (cf. DeCoster, 2009; Hedges & Olkin,1985; Rosenthal, 1991). In the case of t statistics (or pairwise comparisons), we assumed the more conservative two-tailed tests when no information regarding this point was given.

When in a single study measurements are reported that are made both during and after stressor presentation, only the measures obtained during stressor presentation were used. Exceptions to this rule are endocrine measures, since these measures are virtually always measured poststressor due to the larger response latency. For studies in which multiple measurements were performed during stressor presentation (this occurred for 10 studies), we used the values from measurements performed at the time closest to the mean reported measurement time in the overall study sample, which was 16.8 min. In this manner, we aim to include the values representing the measurement time span that is the most consistent with the measurement values in the rest of the study sample.

If the exact values of the needed statistics were not explicitly reported in a study, we derived these values from a plot if available, using the program Plot Digitizer (Department of Physics, University of South Alabama, 2012). In a number of studies, it was not possible to calculate the effect size using the methods described above. In these instances, we contacted the corresponding author with a request for the necessary statistics to calculate the effect size. When it was not possible to retrieve the needed data from the authors in this method, we used a p value of .05 to calculate the effect size for those studies explicitly reporting a significant effect and a g^* value of 0 for studies explicitly reporting that no significant effect was achieved. The approaches outlined here resulted in the calculation of effect sizes for 33 different physiological stress responses reported in a final study sample of 48 studies, originating from 37 articles, given in Table 1. (In this table, only the effect sizes of the physiological responses later used in the analyses are reported. A version of the table containing calculated effect sizes of all 33 physiological variables can be requested from the corresponding author.) Studies not contained in this final sample were discarded because the necessary information to calculate the effect size was not reported in the article and could not be provided by the author.

Statistical Analyses

To gain insight in both the mean effect of a digital game stressor and the effects of specific game and study characteristics on the physiological stress responses included in our study sample, two consecutive analyses were performed. First, the mean effect of a digital game stressor on all physiological signals is determined. Second, a predictor analysis is performed on the physiological signals for which significant variance between effect sizes (heterogeneity) was found and sufficient data is available. For all analyses, we treat the study factor as a random effect; that is, we assume that studies estimate a different effect size since these studies are performed on different populations that have different average effect sizes due to the varying properties between populations. This approach has been argued by several authors to be the norm for social science data, and it allows us to generalize our findings beyond our study sample (A. Field, 2009)

Mean effect size analysis. In this first analysis, we determine the mean effect size of being presented with a digital game stressor on each physiological signal separately, including the significance and heterogeneity (i.e., the variance of this effect across studies) of this mean effect. We used the method proposed by Hedges and colleagues (cf. Hedges & Olkin, 1985; Hedges & Vevea, 1998). This method is applied because it controls Type I errors better than the other prevailing method, by Hunter and Schmidt (2004), in cases

where more than 20 studies are included(Field & Gillett, 2010), such as in our study sample.

This method entails weighing the effect sizes of the different studies by the inverse of their variance together with a between-study variance estimation. By doing so, we take into account both the different sizes of the studies included (by weighing using the inverse of the variance), and the between-study variance of the effect size assumed by our random effects approach. These weighed effect sizes are subsequently used to determine the mean effect size, the significance, and the heterogeneity of the effect size.

Additionally, to gain insight into the potential effects of publication bias (i.e., the tendency to favor publishing certain results over other results), we calculated Rosenthal's fail-safe N. This statistic indicates the number of unpublished studies that needs to exist to turn a significant effect size into a nonsignificant effect size (Rosenthal, 1979).

Predictor analyses. Subsequently, in order to assess whether individual study or game characteristics are related to any significant variation we find in physiological stress responses, we performed predictor analyses between the effect sizes and the game and study characteristics for each of these physiological variables separately. In performing the analyses separately per physiological variable, it is ensured that the assumption of independence between effect sizes within one analysis is maintained, because no study will deliver more than one effect size to each analysis. Two predictor analyses have been performed per physiological signal. The first analysis contains both game and study characteristics as potential predictors, and the second contains only game characteristics, to gain insight specifically into the effect of game characteristics on the different physiological variables. As the aim of this meta-analysis is to gain insight in the effects of digital game characteristics, no separate analysis is performed on solely study characteristics, as this is not within the aim of the study.

We used the algorithms provided by Lipsey and Wilson (2000). This approach consists of constructing weighted linear regression models in which the effect sizes are weighted by the inverse of their variance together with a between-study variance estimation, in the same way as was done in the previous mean effect size analysis (for a detailed treatment of this approach, see DeCoster, 2009; Field & Gillett, 2010; Hedges & Olkin, 1985). For each analysis, we fitted a separate regression model in which the respective physiological variable was the dependent variable. We first transformed our categorical predictor variables with more than two levels (i.e., the game characteristics input modalities and game type) into several categorical variables with only two levels (i.e., dummy coding), in order to make the data suitable for use in the analyses. This means that for each of these predictor variables the different levels of the variable are now being used as separate variables and can be reported as such in the coming analyses. For example, instead of using and reporting the predictor game type, we can now specifically report the effect of the fast puzzle game type or the racing/sports game type variable.

All regression models were constructed in an initial hierarchical manner by initiating the model with predictors having the strongest theoretical basis for being predictors of the effect size, the characteristics aversive stimuli, and realism. This was based on the fact that, in the study sample, the subset containing studies with aversive stimuli and the subset with studies containing high realism values had the highest effect size.

After this initiation of the model, predictors were added to the model in a forward stepwise manner, using two rules. This forward method of model construction was preferred over a backward stepwise or forced entry method since these methods would initialize the regression models with 10 predictors, which would be too many predictors for the respective 40, 28, and 28 studies in the models, potentially causing overfitting of the data.

The first rule that was applied was to eliminate the most nonsignificant (i.e., the least contributing) predictor from the model if there was a nonsignificant predictor and as long as this would leave at least one predictor in the model. If this rendered any of the remaining significant predictors to become insignificant or provided a significant reduction in the model fit, the predictor was restored to the model. The second rule, which was executed if the first rule could not be applied, was to add a new predictor, not previously used with this set of predictors, with the highest semipartial correlation to the effect size. If this provided a significant increase in the fit of the model, the predictor was added to the model; otherwise, the predictor with the next highest semipartial correlation was tried.

Suppressor effects (i.e., the phenomenon in which a predictor has a significant effect only when another predictor is held constant) are known to be more prevalent in the forward stepwise method employed in our model construction (A. Field, 2009). In order to account for these effects, we constructed each regression model using two different levels of significance for the second rule, $\alpha = .05$ and $\alpha = .1$. The α of .1 allowed for faster inclusion and inclusion of more predictors, reducing the risk of Type II errors, associated with suppressor effects. When these two approaches resulted in different models, we used the model containing the most significant predictors. In certain cases, the two models derived using different levels of significance included the same significant predictors, although one had additional nonsignificant predictors. In those instances, we used the model without the additional nonsignificant predictors.

Results

Mean Effect Size Analysis

The results of the mean effect size analysis for physiological variables used in at least four studies are given in Table 2. (The complete table, containing the mean effect size analysis for all measured physiological responses, can be requested from the corresponding author.) These results show that four physiological stress responses have both significant mean effect sizes and significant heterogeneity of effect sizes ($p < .05$, see columns pe and p_h of Table 2, respectively) across the study sample. To gain insight in the sources of this heterogeneity, predictor analyses are performed on physiological signals of which more than 10 studies provided effect sizes (second column, Table 2). This is done to prevent deriving false relations due to a lack of studies (A. Field, 2009). Thus, for HR, SBP, and DBP (respectively, 40, 28, and 28 effect sizes), predictor analyses were performed.

Furthermore, the values of the fail-safe N for these three signals (last column, Table 2) fulfill the criterion proposed by Rosenthal stating that an N higher than $5k + 10$ (with k being the number of studies) reflects a reliable meta-analytic effect (Rosenthal, 1979): for HR, DBP, and SBP, the respective fail safe Ns (4,684, 7,360, and 7,426) are well above the criterion values of 210, 150, and 150, respectively.

Predictor Analyses

Predictor analyses are performed on HR, SBP, and DBP. Given the results described in this section, we also performed additional analyses, described in the subsection Subanalyses.

Main analyses. In this main analysis, regression models were constructed according to the rules described in the Statistical Analysis section above. The process of constructing these models is given in Table A2. The regression models using both game and study characteristics are based on 40, 26, and 26 studies, and the models based on only game characteristics are based on 39, 27, and 27 studies for HR, DBP, and SBP, respectively. The total variance that is explained by these models when using both game and study characteristics is 41% of the total variance ($R^2 = .412$) for HR, the model for DBP explains 32% of the total variance ($R^2 = .319$), and the model for SBP explains 34% of the total variance ($R^2 = .336$). When using solely game characteristics, the regression model for HR explains 40% of the total variance ($R^2 = .405$), the model for DBP explains 13% of the total variance ($R^2 = .125$), and for SBP 11% of the total variance ($R^2 = .105$) is explained. These results are visualized in the first row of Figure 1.

For HR, in both models, game music, fast puzzle game type, and realism were significant predictors. For DBP, racing/sports game type is a significant predictor in both models, and in the model using also study characteristics, reliability of effect sizes was an additional significant predictor. For SBP, the model using only game characteristics contained only the characteristic aversive stimuli. The model using both types of characteristics contained the predictors reliability of effect size, racing/sports game type, and stressor duration. These models, together with the corresponding descriptive, are provided in Table 3.

All characteristics are visualized in Figure 2, with the lines ending with diamonds depicting the characteristics from this main analysis. In this figure, the lines indicate the range of values a specific characteristic can have and the predicted change of effect size in HR or blood pressure associated with this range of values.

Additionally, we checked the assumptions that need to be met for the results of this analysis to be generalizable beyond our study sample, checked for any outliers with the derived models, and controlled for influential cases. The tests performed and the test statistics are given in Table A3 for the sake of readability. We found all assumptions had been met for all models, except for the models for DBP and SBP using only game characteristics—these both did not meet the assumption of independence of errors. No outliers were observed for any of the models. We controlled for influential cases by examining the differences between the derived models and the models excluding their respective influential cases. For DBP, no differences were found when excluding these influential cases. For the derived model for SBP using both game and study characteristics, it was found that excluding influential cases rendered the reliability of effect size predictor no longer significant ($p = .08$). Also, in the model using only game characteristics, the aversive stimuli predictor was rendered no longer significant ($p = .663$) when excluding influential cases. Finally, the derived model for HR using only game characteristics was found to render the aversive stimuli predictor significant ($p = .04$) when excluding influential cases, whereas this was not the case with the influential case included.

Furthermore, there is a relatively high number of studies issued from one research group, being the studies from Miller and colleagues (see the eight studies in Table 1). We therefore performed t tests for the continuous variables and χ^2 goodness of fit tests to see

Table 2. *Overview of Weighted Mean Effect Analysis for all Coded Physiological Variables Used in at Least Four Studies*

Physiological stress response	Number of studies	Number of participants	Effect size								Heterogeneity		
			min g^*	max g^*	mean g^*	-95% CI	$+95\%$ CI	SE g^*	Z	pe	Q	df	p_h
HR	40	1,249	−0.701	2.639	0.7296	0.5357	0.9236	0.099	7.3723	< .00005	203.578	39	< .00005
DBP	28	2,671	−0.256	3.074	0.9381	0.7615	1.1147	0.0901	10.4109	< .00005	156.0734	27	< .00005
SBP	28	2,671	−0.385	3.56	0.9233	0.7482	1.0984	0.0893	10.3355	< .00005	152.5993	27	< .00005
FBF	5	150	−0.134	2.535	0.8796	0.0844	1.6784	0.4057	2.168	.0302	39.4947	4	< .00005
RR	5	99	−1.214	1.721	0.9371	−0.0303	1.9044	0.4936	1.8986	.0576	37.4191	4	< .00005
FVR	5	150	−0.71	0.033	−0.3610	−0.6295	−0.0925	0.137	−2.6349	.0084	5.2437	4	.2632
CO	4	185	0.186	0.652	0.3815	0.1755	0.5876	0.1051	3.6293	.0003	2.9974	3	.392
DBVP	4	80	−1.441	−0.167	−0.8437	−1.3781	−0.3094	0.2726	−3.0947	.002	7.3144	3	.0625

Note. Min, max, mean g^* = minimum, maximum, and mean effect size found in all studies for this variable, respectively; -95% CI, +95% *CI, SE, Z, pe* = confidence interval, standard error corresponding Z value, corresponding p value; Q, df, p_h = indicate the heterogeneity, corresponding degrees of freedom, and corresponding p value; CO = cardiac output; DBP = diastolic blood pressure; DBVP = digital blood volume pulse; FBF = forearm blood flow; FVR = forearm vascular resistance; HR = heart rate; RR = respiration rate; SBP = systolic blood pressure.

if there are significant differences in any of the predictor variables or the dependent variable when excluding these studies from the dataset. These tests were nonsignificant for all variables.

Subanalyses. Based on the results in the main predictor analyses, we performed an additional analysis on a subset of the data. This analysis is based on the finding that the newly constructed study characteristic reliability of effect sizes is a significant predictor of DBP and SBP effect sizes. We therefore constructed an additional regression model for both the DBP and SBP to gain more insight in the moderating effect of game and study characteristics in studies scoring high on our reliability of effect size scale (i.e., in which less interference is to be expected due to higher methodological quality). Therefore, we excluded the quartile of studies exhibiting the lowest values for this characteristic. The remaining subset corresponds to 21 studies, all having a value of 10 or higher for the reliability of effect sizes characteristic. The models are initiated with the same predictors and constructed using the same rules used in the main predictor analyses. The process of constructing these models is given in Table A4. The variance that is explained when using both game and study characteristics is 57% of the total variance ($R^2 = .5680$) for DBP and 49% of the total variance ($R^2 = .4924$) for SBP. When using solely game characteristics as predictors, the derived regression models explain 23% of the total variance ($R^2 = .2261$) in DBP change and 22% of the total variance ($R^2 = .2183$) in SBP change. These results are visualized in the second row of Figure 1.

For DBP, the model using only game characteristics contains one significant predictor, racing/sports game type. The model using both game and study characteristics also contains this significant predictor, together with the reliability of effect size characteristic as significant predictor. SBP has three significant predictors when using both types of characteristics: reliability of effect sizes, stressor duration, and fast puzzle game type. The model using solely game predictors contains only aversive stimuli as significant predictor. These predictors, together with corresponding descriptive, are given in Table 4.

These characteristics are also visualized in Figure 2, with the lines ending with circles depicting the characteristics forming this subanalysis.

With regard to the assumptions that need to be met in order for the results to be generalizable beyond our study sample, we found that all assumptions for all derived models had been met. Furthermore, no outliers and no differences when excluding influential

cases were observed for any of the models in both subanalyses. Detailed descriptions of the performed tests and results thereof for determining the fulfillment of assumptions, detecting outliers, and controlling for influential cases can be found in Table A3.

An overview of all predictors is given in Table 5, with the coefficients averaged over the results of the different analyses.

Discussion

In this meta-analysis, we analyzed 48 studies that measured physiological stress responses to digital game stressors. In the mean effect size analysis, we observed significant effect sizes for 13 out of the 33 physiological variables in our study sample, with effect sizes up to 0.94 (Table 2), which is considered a large effect size (Cohen, 1988). This indicates that, consistent with the literature, digital games elicit a broad spectrum of physiological stress responses of considerable magnitude. We also found significant heterogeneity in the effect sizes of six of the physiological responses, indicating more variance than could be expected due to chance. Using predictor analyses on three of these physiological responses for which sufficient studies were available, being HR, DBP, and SBP, we found specific sets of game and study characteristics moderating these different physiological responses. These sets explain between 11% and 57% of the variance in the physiological responses and contain in total six significant predictors: game type, aversive stimuli, realism, game music, reliability of effect size, and stressor
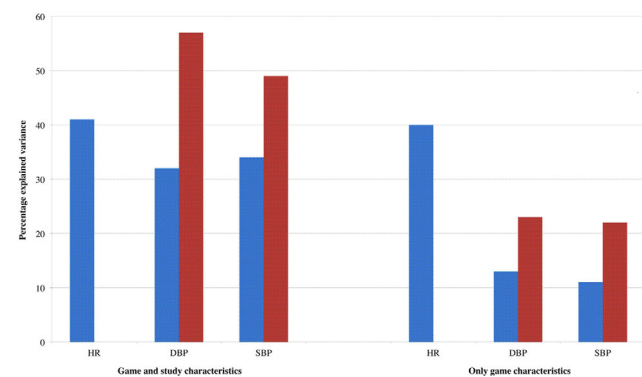

Figure 1. Percentage of variance in physiological responses that is explained by solely game or both game and study characteristics in the main and subanalyses. Blue bars indicate main analyses, red bars subanalyses.

Table 3. *Overview of Main Analysis Final Regression Models for each Physiological Variable*

Physiological variable	Regression model using both game and study characteristics					Regression model using only game characteristics				
	Predictors	β	Z	P	R^2	Predictors	β	Z	P	R^2
Heart rate	Constant	.0000	6.1636	.0000	.412	Constant	.0000	6.6045	.0000	.405
	Game music	−.5081	−3.4985	.0005		Game music	−0.5234	−3.6036	.0003	
	FPGT	−.2847	−2.0361	.0417		FPGT	−0.3696	−2.5755	.0100	
	Realism	−.4315	−2.7039	.0069		Realism	−0.4917	−2.7545	.0059	
	Stressor duration	.1717	1.1140	.2653		Aversive stimuli	0.2673	1.6276	.1036	
Diastolic blood pressure	Constant	.0000	4.3798	.0000	.319	Constant	0.0000	6.8503	.0000	.125
	Reliability of effect sizes	−.5832	−3.5947	.0003		RSGT	0.3534	2.6990	.0070	
	RSGT	.5865	3.9623	.0001						
	Stressor duration	.1430	1.0004	.3171						
Systolic blood pressure	Constant	.0000	3.7029	.0002	.336	Constant	0.0000	11.1772	.0000	.105
	Reliability of effect sizes	−.4058	−2.5143	.0119		Aversive stimuli	−0.3241	−2.5250	.0116	
	RSGT	.4049	2.7531	.0059						
	Stressor duration	−.2886	−2.0425	.0411						

Note. Overview of main analysis final regression models for each physiological variable, using either both game and study characteristics or solely game characteristics as predictors. For each model the explained variance R^2 is given, and for each predictor and constant of the respective regression models the following parameters are given: the standardized coefficient (ß), expressing the change in the respective physiological variable in standard deviations given the change of one standard deviation in the corresponding predictor, the Z-value, and the corresponding significance, p. FPGT = Fast puzzle game type; RSGT = Racing/sports game type.

duration. Below, we first discuss these significant predictors separately, after which we examine the characteristics that were found to be nonsignificant. Finally, we integrate the findings from all these characteristics into a new model, depicting the found moderating relations and the hypothesized underlying processes.

Game Type

For both HR and DBP, game type is a significant predictor in all analyses. For SBP, this holds for the main and subanalyses using both game and study characteristics (Figure 2). Furthermore, a clear distinction can be seen regarding what type of game moder-

ates HR and what type of game moderates DBP: puzzle game type significantly moderates the HR response, whereas the racing/sports game type moderates the DBP stress response. For SBP, a more diffuse image arises: in the main analysis, the racing/sports game type moderates the response whereas in the subanalysis, the fast puzzle game type does this.

Regarding the direction of the moderation of the two game types, a clear image can be observed: in all analyses, the racing/sports game type has a positive moderating function; that is, the increase in the respective physiological response is strengthened by using a racing/sports game type compared to the other game types included in our analyses, such as a fighting game. The
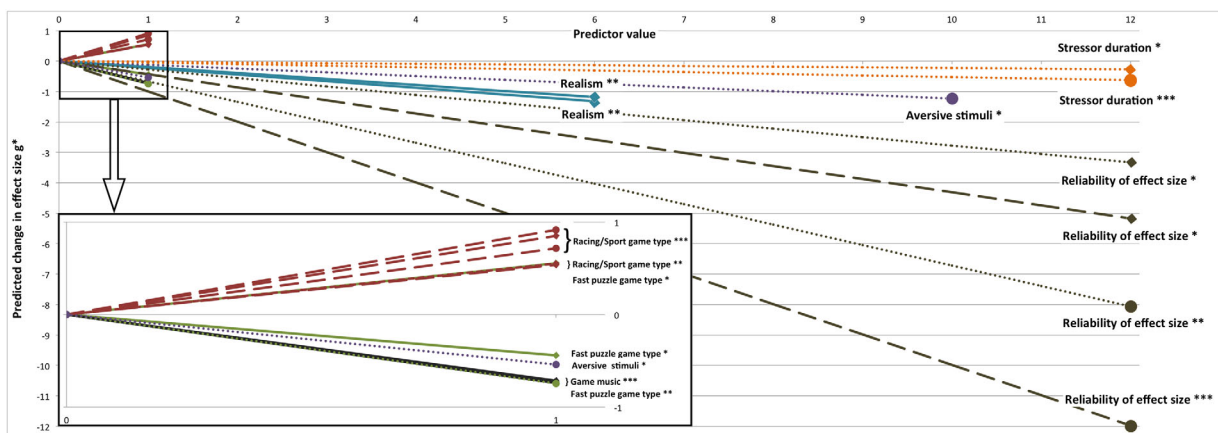

Figure 2. Visualization of all significant predictors in all regression models. Solid lines present predictors for HR, dashed lines indicate predictors for DBP, and dotted lines indicate predictors for SBP. Values on *x*-axis represent values of the respective predictor, and length of the lines indicates the possible range of values the predictor can have (except for stressor duration, which, in theory, can take any positive value). See Method for more information on the possible values for the different predictors. For consistency, discrete variables are also depicted using a line, even though it is not possible for these characteristics to have any other value than positive integers. Lines ending in a diamond represent predictors derived during the main analyses, lines ending in a circle depict predictors derived during the subanalyses. Values on *y*-axis represent the expected change in effect size for the given value of a predictor on the *x*-axis. The inset shows the area of the graph around the origin in greater detail to allow better insight in the characteristics depicted in this area. *p< .05; **p< .01; ***p< .001.

Table 4. *Overview of Derived Regression Models in Subanalysis*

Physiological variable	Regression model using both game and study characteristics					Regression model using only game characteristics				
	Predictors	β	Z	P	R^2	Predictors	β	Z	P	R^2
Diastolic blood pressure	Constant	0.0000	4.5419	.0000	.568	Constant	0.0000	3.3052	.0009	.226
	Reliability of effect sizes	−0.6480	−4.2671	.0000		RSGT	.4755	2.5247	.0116	
	RSGT	0.6366	4.1918	.0000						
Systolic blood pressure	Constant	0.0000	3.4980	.0005	.492	Constant	.0000	6.1817	.000	.218
	Reliability of effect sizes	−0.4513	−2.8411	.0045		Aversive stimuli	−.4378	−2.2505	.0244	
	Stressor duration	−0.6229	−3.7163	.0002		FPGT	−.3647	−1.8751	.0608	
	FPGT	−0.5318	−3.0708	0.0021						

Note. Overview of derived regression models in sub-analysis, constructed excluding quartile of studies with lowest reliability of effect size values for diastolic and systolic blood pressure, using either both game and study characteristics or solely game characteristics as predictors. For each model the explained variance R^2 is given, and for each predictor and constant of the respective regression models the following parameters are given: the standardized coefficient (ß), expressing the change in the respective physiological variable in standard deviations given the change of one standard deviation in the corresponding predictor, the Z-value, and the corresponding significance, p.
FPGT = Fast puzzle game type; RSGT = Racing/sports game type.

magnitude of this moderation lies between 0.55 and 0.91, indicating that the stress response is between 0.55 and 0.91 standard deviations higher when using a racing/sports game compared to other game types. In one of the analyses, the assumption of independence of errors was not met, indicating we should be careful with generalizing these findings beyond our study sample. However, the same positive moderating function for the racing/sports game type characteristic is found in four other analyses, rendering it acceptable to generalize this finding beyond our study sample. The moderation direction of the fast puzzle game type is precisely the other way around: it decreases the physiological response compared to other digital game types in our analysis, with a magnitude differing from −0.44 to −0.74. An exception to this observation can be observed in the HR predictor analysis using only game characteristics. Here, the fast puzzle game type has a coefficient of 0.55. However, the correlation coefficient of this predictor with the effect size is negative ($r = -.12$). Given this contradiction between the direction of the correlation and predictor coefficient, we treat this predictor as a suppressor variable (i.e., a variable that, when introduced to the regression model, increases the variance explained by the complete model, but does so by either accounting for the residual error in the model or by enhancing the variance explained by other

predictors). In general, these variables are then not interpreted (Thompson & Levine, 1997).

This difference between the directions of both game types can most likely be explained by the type of mental activity involved and the required preparation for motor activity. Arguably, fast puzzle games in general involve more mental activity, whereas racing/sports games involve more preparation for motor activity. Although different findings exist in the literature regarding the effects of mental activity on HR responses, there are indications that HR drops with increased visual processing and attention (Codispoti, Bradley, & Lang, 2001; Graham, 1992), which is consistent with our findings. Additionally, preparation for motor activity has been found to increase blood pressure (Lovallo, Stress, & Health, 2005), which is also consistent with our findings.

Aversive Stimuli

Aversive stimuli are a significant negative predictor of the SBP in both the main and subanalysis when using only game characteristics (Figure 2), with respective coefficients of -0.12 and −0.16. (Table 3 and Table 4).

The set of 28 studies used for deriving the main and subanalysis models for SBP contained only two studies with aversive stimuli.

Table 5. *Significant Moderating Relations Between Game and Study Characteristics and Physiological Stress Responses*

Stressor	Physiological stress response		
	Heart rate	Diastolic blood pressure	Systolic blood pressure
Mean effect digital game stressor	0.73	0.94	0.92
Game type—fast puzzle	−0.44[a]	0	−0.74
Game type—racing/sports	0	0.75	0.55
Aversive stimuli	0	0	−0.14
Realism	−0.21	0	0
Game music	−0.73	0	0
Reliability of effect size	0	−0.71	−0.47
Stressor duration	0	0	−0.04

Note. Values are given in $g*$ for the digital game stressor and in absolute change of $g*$ for each moderating characteristic, per unit of the given characteristic, with 0 indicating that no significant moderation was found. Values are averaged over the results of different analyses because the direction is the same for the averaged values.
[a]Value of 0.55 for the main analysis using only game characteristics is omitted, as this is determined as a suppressor effect; see discussion of this characteristic.

This is also reflected in the fact that the main analysis model failed to meet several assumptions regarding influential cases (see Table A3). In line with recommendations made by Field & Gillett (2010) regarding the interpretation of such underrepresented levels of predictors, we restrain from further interpretation of this finding here, since it is not clear whether this is a true effect or a spurious relation induced by underrepresented levels of aversive stimuli.

Realism

The amount of realism negatively moderates the HR stress response. In both main analyses, the coefficients of the realism predictor are close to one another: −0.2 and −0.22, respectively, indicating a dampening effect on the HR stress response with an average of −0.21 standard deviation for each point of realism on the scale used in our analysis.

The potential effects of this moderation are quite large as the maximal difference between realism values encountered in our study sample is 4.67 (range between 1 and 5.67, see Table 1.). This entails that the potential difference in effect size measured between stressors due to differing realism is 1.03 *SD* (0.22 * 4.67), which is a large effect on itself (Cohen, 1988) and can therefore easily be the difference between an encountered small or large effect size.

The finding of a dampening effect of realism on the HR response is not consistent with the literature, where in general an increase of realism is correlated with an increased physiological stress response (Barlett & Rodeheffer, 2009; Ivory & Kalyanaraman, 2007). However, our finding is reconfirmed when the simple regression coefficient between realism and the HR stress response in our study sample is determined, being -0.337. This contradictory finding could be due to the fact that our definition of this characteristic was a composite of three factors, whereas most studies use a definition based on either one of these factors. The definitions in themselves do not vary substantially from definitions previously used by other authors, as we based our definitions on existing ones. A possible explanation could be that an interaction exists between the used factors, which results in a moderating effect opposite to what is observed in the literature. For example, it could be that there is an interaction between year of release and the presence of a human avatar, in the sense that in older games the presence of a human avatar did not add, but perhaps even subtracted to the experienced realism. This could be due to the fact that graphic abilities of digital games in that time did not allow for a realistic depiction of a human avatar, whereas the depiction of an abstract avatar could be perceived as more realistic because these do not have a counterpart in the real world to compare the graphic depiction against. Another possible explanation could be that more realistic digital game stressors require more visual processing and attention, resulting in a decrease of HR, as discussed earlier.

Game Music

Game music is a significant negative predictor in both main analyses of the HR stress response. The magnitude of the moderation is in both main analyses consistent with a coefficient of −0.71 and −0.74, when using both game and study characteristics or when using only game characteristics, respectively. This means that when presenting a digital game stressor with accompanying music the HR stress response is decreased with 0.71 or 0.74 standard deviation compared to presenting the same stressor (in terms of our characteristics) without game music. This moderation potentially has a large impact, as a difference of 0.74 in a measured effect size

due to the absence or presence of game music can easily change a small effect size to a large effect, or vice versa.

This finding of decreased response when including music is not completely consistent with the literature, where the presence of music during playing of digital games is mostly linked to positive effects on the HR change (Nyklicek et al., 1997). However, we were not able to take into account the type of music presented in our analysis (as it was not possible to code this consistently across studies). Recently, a study by Bernardi et al. (2009) found consistent relations between cardiovascular responses and music profiles such as crescendos and rhythmic phrases. Therefore, additional research is needed to gain insight into the effect of music in the context of a digital game stressor across several studies, using a form of objective music analysis such as used by Bernardi et al. (2009). An alternative explanation could be that the addition of game music to a digital game stressor requires more auditory processing and attention, resulting in a decrease of HR, as discussed earlier.

Reliability of Effect Size

The new characteristic introduced in this analysis, reliability of effect size, emerges as a significant negative predictor of both the SBP and DBP response. This finding is consistent across the different analyses and entails that when a study scores lower on the reliability of effect sizes scale (i.e., fulfills fewer of the factors stated at the introduction of this characteristic) an overestimation of the DBP and SBP stress response is to be expected. This observed moderation has a coefficient of −0.43 for DBP and −0.28 for SBP in the main analyses and −1 and −0.67 in the subanalyses (Table 3 and Table 4). Here, the same effect found in other characteristics can be observed: when examining studies with high scores for the reliability of effect size characteristic (i.e., the subset of studies used in the main analysis), the moderating relations found on the complete set of studies are enhanced and significance is increased (significance is increased from .0001 to < .00005 for DBP and from .0119 to .0045 for SBP, see Table 4).

The implications of not fulfilling the requirements are large: for every requirement not fulfilled in the reliability of effect size scale, the DBP and SBP are predicted to be overestimated with 0.43 and 0.28 standard deviation. The maximal difference in values for this characteristic in our study sample is 3 (range between 8 and 11) for studies included in the main analyses and 1 (range between 10 and 11) for studies included in the subanalyses, for both DBP and SBP (see Table 1. and Figure 2). This entails that the potential overestimation in measured effect sizes due to not meeting the requirements enclosed in the reliability of effect size scale is estimated at a maximum of 1.29 *SD* for DBP and 0.84 *SD* for SBP, for studies scoring below 10 points on this scale. For studies scoring above 10 points on this scale, the potential overestimation is estimated at 1 *SD* for DBP and 0.67 *SD* for SBP. This is the largest moderation found in this analysis and can potentially alter effect sizes from small to large and vice versa.

Of all coded factors determining the reliability of effect size value of the studies measuring DBP and SBP, three factors account for 89% of the value of this characteristic. These factors are stressor substance refrainment, the presence of preceding tasks, and stressor onset timing, The observed overestimation is indeed to be expected with respect to these influential factors.

Regarding the stressor substance refrainment factor, several studies (e.g., Smits, Temme, & Thien, 1993) have shown increased blood pressure stress responses under the influence of nicotine, caffeine, or a combination of both. This effect is more confounded

for the HR response, which was found to increase after nicotine intake, decrease after caffeine intake, and not change significantly when both substances are present. These findings are consistent with our prediction of overestimated effect sizes for blood pressure for studies in which participants are not refrained from these substances prior to measurement, and our finding of no significant moderation of the heart rate response.

Furthermore, with respect to the preceding tasks factor, increased blood pressure reactivity is expected due to both prolonged psychological strain on the participant and the resulting increased physiological responses to this strain. Most notably, neuroendocrine responses such as cortisol, which, through various mechanisms, results in increased vasoconstriction (Whitworth, Mangos, & Kel, 2000), will add to higher blood pressure values. The fact that cortisol has a delayed release into the cardiovascular system (Dickerson & Kemeny, 2004) also adds in potentially creating larger blood pressure responses in later tasks. Finally, stressor onset timing is expected to act as a confounder, either increasing or decreasing the effect size, depending on the type of activity performed between baseline measurement and stressor onset.

Stressor Duration

Stressor duration is a significant negative predictor in both the main and subanalysis for SBP (Figure 2). In the subanalysis, we find that stressor duration moderates the increase in SBP with a coefficient of -0.05 (Table 4), meaning that the longer a digital game stressor is presented, the more the general increase of SBP in response to a digital game stressor will be dampened, with a predicted decrease of the effect size of 0.05 per minute. As the stressor duration varies between 3 and 30 min in our study sample, this implies that a maximal difference of 1.35 SD (27 * 0.05) in measured effect size can occur between studies due to differing stressor durations.

We expect this observed dampening effect to be due to two different causes. First, the response to a stressor (assumed that the intensity is kept constant) becomes less when the participant becomes habituated to the stimuli over time. Second, it is expected that prolonged elevated blood pressures are brought down through several homeostatic systems, such as the autonomic nervous system and the baroreflex (Berne & Levy, 2001).

Besides the significant predictors, we also coded four characteristics that were not found to be significant predictors of the physiological stress response: input modalities, additional external stressors, age, and sex.

With regard to input modalities, we found no indication that input modalities have influence on the observed physiological response. As far as we are aware, this has not yet been investigated in the literature. Based on this result and the absence of relevant research in the literature, we hypothesize that it most likely does not matter what type of input controller is used when utilizing digital game stressors, as long as some threshold functionality and comfort is provided.

Concerning the nonsignificance of additional external stressors, we examined the four different factors that comprise this characteristic and investigate the possibilities why these factors result in a nonsignificant compound characteristic.

The first factor is the inclusion of the stressor in a task battery. This factor could have negative effects on measured effect size by raising baseline measurements, effectively reducing the effect between these measurements and response measurements, hereby countering the potential positive effects of

the other factors and reducing the overall expected positive moderation of this characteristic.

The use of social-evaluative threat, incentives or shocks, or loud noise, the other three factors, are also expected to show increased responses in the measured physiological responses. The fact that these results were not found could be because we do not have enough studies to include these additional predictors in the currently constructed regression models.

With respect to the age and sex, we would expect to find these characteristics as significant predictors of physiological stress responses, based on literature utilizing acute psychological stressors (e.g., Hastrup & Light 1984; Matthews & Stoney, 1988; Stoney et al., 1988). We did not, however, find these characteristics to be significant predictors in our analyses.

We hypothesize this to be due to intrinsic differences between digital game stressors and the acute psychological stressors used in the referenced literature, which utilize tasks such as serial subtraction, reaction time, or speech tasks. To gain more insight into this matter, we examined the studies in our study sample that specifically investigated sex differences, and compared the found physiological responses to digital game stressors with the responses to the different acute psychological stressors found in the literature.

Three studies were found (Lawler et al., 1995; Markovitz, Raczynski, Wallace, Chettur, & Chesney, 1998; Tischenkel et al., 1989) that investigated sex differences. Regarding blood pressure, the results are comparable with the findings for other psychological stressors. In two studies of our study sample, males show increased blood pressure responsivity (Tischenkel et al.; Lawler et al., although the latter only found this for DBP), which corresponds with what is found in the literature for other types of acute psychological stressors. However, in the third study by Markovitz et al., there was no difference found between the sexes. Regarding HR, Tischenkel et al. found no differences, and Lawler et al. found an increased HR responsivity for males. This is in contrast with HR responses found for other acute psychological stressors, for example, Stoney et al. (1988), who found that females showed increased HR responses to a speech task compared to males. This same approach was not possible for age, as there were no studies in our study sample that investigated the moderating effects of this characteristic within one study.

Therefore, we cannot explain the nonsignificance of age and sex characteristics based on intrinsic differences between digital game stressors and other acute psychological stressors, as only for HR differences could they be observed, and then only based on a small number of studies.

Given these insights, we expect the cause of the nonsignificance of age and sex to be methodological. We hypothesize that the moderating effect of these characteristics is not large enough compared to the other characteristics to be included during construction of the regression models. During this construction, characteristics with small moderating effects can fail to significantly increase the explained variance of the model, resulting in the absence of these characteristics in the regression models.

When we take the discussion on the separate characteristics together, we can come to several interesting insights and the resulting implications.

First, the exact three game characteristics that were found to be significant predictors in the main analysis using only game characteristics are also found in the constructed models using both game and study characteristics. Even more, the characteristics moderate the exact same set of physiological responses with magnitudes highly comparable to the magnitudes observed
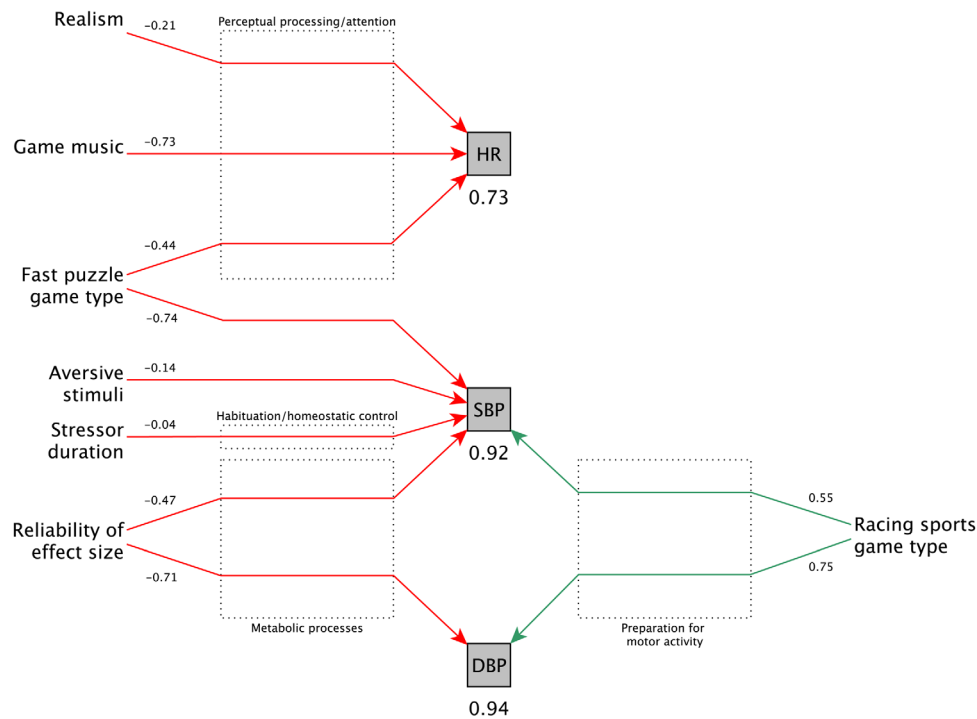
Figure 3. Overview of significant moderating relations between digital game and study characteristics and physiological responses, including the hypothesized processes underlying these relations. Green arrows indicate positive relations, whereas red arrows indicate negative moderations. Values given at arrows indicate the moderating effect, i.e., the expected change, expressed in g^*, when the given characteristic increases one point (points differ per characteristic, as defined in Method). The values given at the physiological responses are the mean effects on these responses when presented with a digital game, expressed in g^*. Dotted rectangles indicate the cognitive and physiological processes underlying the moderations crossing through these areas.

in the analyses using both type of characteristics. This strengthens the findings of both analyses and indicates the consistency of the found relations between digital game stressor characteristics and physiological responses.

Second, the models that were found in the main and subanalyses for the different physiological responses are both highly comparable within one response, but clearly different between responses, most notably between HR and blood pressure. This indicates that moderating relations between these characteristics and physiological responses are highly specific for each physiological response and consistent within each physiological response.

Third, by adding study characteristics to the predictor analyses, a sharp increase in explained variance of the physiological responses can be observed. The mean explained variance in the regression models using only game characteristics is 22% of the total variance observed in the physiological stress responses. This percentage increases to an average of 43% when adding study characteristics as potential predictors, with a maximal explained variance of 57%. This indicates that study characteristics (i.e., study methodology) have an important influence on the measured stress response, comparable to game characteristics.

Of these study characteristics, the newly introduced reliability of effect size characteristic has the highest potential moderating effect. We found that significant predictors for the whole study sample are also predictors for the subset with studies scoring the highest on this scale. In this subset, the significance and magnitude of the moderation have increased, hereby also increasing the explained variance, which is on average 53% in the analyses utilizing game and study characteristics. This confirms that the newly developed reliability of effect size characteristic indeed has an effect when applied, and

indicates that, when studies utilize the correct methodology (i.e., scoring higher on the reliability of effect size scale), the impact of stressor and study characteristics is larger, hereby granting the researcher more control over the elicited stress response and more insight in underlying processes and observed responses. Furthermore, it reconfirms the findings of the main analyses.

Fourth, the presence of such specific and consistent moderating functions suggests that the same processes underlying these moderations were activated throughout the different analyses. When we take the potential underlying processes for the different characteristics together, a distinct grouping of moderations emerges: moderations are expected to be underpinned by either perceptual processing and attention, homeostatic control, preparation of motor activity, or metabolic processes. This grouping is visualized in Figure 3: perceptual processes are expected to underlie the negative (inhibitory) relation between visual and auditory input (game music, realism, and fast puzzle game type) and HR responsivity (Bernardi et al., 2009; Codispoti, Bradley, & Lang, 2001; Graham, 1992). Preparation of motor activity is hypothesized to underlie the positive (excitatory) relation between game type and blood pressure (Bernardi et al., 2009; Lovallo, Stress, & Health, 2005). Homeostatic control is proposed to underlie the relation between stressor duration and SBP (Berne & Levy, 2001), and metabolic processes are influenced by endogenous and exogenous chemicals, which is expected to underlie the negative relation between the methodological factors and blood pressure (Dickerson & Kemeny, 2004; Smits et al., 1993; Whitworth et al., 2000)

Additionally, the specificity of these moderating relations support a concept in which there is more response specificity of

cardiovascular stress responses than previously assumed. This calls for further research into the exact properties of this specificity.

Fifth, when we take the high explained variance of the derived regression models, with up to 57% of explained variance in physiological responses, into account, this finding supports the notion that digital game stressors are not stressors by virtue of being a digital game. In contrast, our findings indicate that a digital game stressor derives its stressor function from the different characteristics it is composed of, and the methodology through which it is applied. This supports a concept in which digital game stressors should be viewed as a container construct in which several specific stressor types and cognitive tasks, such as aversiveness and induced cognitive load, can be placed, in order to create the complex stressors that we refer to as digital game stressors.

The characteristics found in this analysis to moderate the physiological stress response are not exotic or rarely encountered properties of digital game stressors and studies, but occur frequently in our study sample and show substantial variation across studies. However, these characteristics are currently not recognized for having such large and specific effects on the measured physiological response. Therefore, our findings may have two implications.

The first implication is that the observed specific and heterogeneous results indicate that there is a need for a more standardized version of digital game stressor types, analogous to what, for example, the Trier Social Stress Test (Kirschbaum, Pirke, & Hellhammer, 1993) is for public speaking stressors. In the absence of such a standard, the inevitable differences between digital game stressors and methodology used in various studies will cause differences between observed results, troubling the synthesis of research results and the understanding of underlying stress response systems. Efforts in this direction have been undertaken by van der Vijgh, Beun, van Rood, & Werkhoven (2014), who have proposed a standardized digital game stressor based on findings from this meta-analysis.

The second implication is that, in absence of such a standardized stressor, the results from this analysis can be used as a guideline in designing or utilizing different digital game stressors and corresponding methodology, or when comparing results from studies utilizing this type of stressor. When designing a stressor for use in research, characteristics can be implemented in the stressor based on the found moderating effect of the characteristics on the response that will be measured. For example, when designing a digital game stressor for inducing blood pressure stress responses, a racing/sports game type would be preferred over other game types, as this game type increases both DBP and SBP. All specifics regarding these moderating functions that can be used to guide the design process in more detail are given in Table 5.

Furthermore, when designing a physiological stress experiment using digital games, attention must be paid to study methodology in order to determine the correct stress responses to the presented digital game stressors, allowing for more insight in the underlying processes and observed responses. Regarding this methodology, two study characteristics have been found to moderate the measured physiological response, being the new reliability of effect size scale, introduced in this study, and the stressor duration.

Of these characteristics, the reliability of effect size characteristic has the largest moderating function. The findings indicate that, when a lower score on this scale is achieved, an overestimation of the observed stress response is predicted. Three factors have been identified that account for the largest part of the moderating function, being the stressor onset timing, potential preceding tasks, and substance refrainment. In terms of methodology, the findings suggest keeping the stressor onset timing at a minimum, including no

preceding tasks, and refraining participants from nicotine and/or caffeine prior to the experiment.

The second moderating study characteristic is the stressor duration, which entails a dampening effect on the stress response with prolonged presentation of the stressor. This implicates that the length of stressor presentation should be taken into account in the methodology design

In line with these methodological implications, it follows that simple one-to-one comparisons and interpretations of results between physiological stress studies are in most cases not possible; the characteristics described above need to be taken into account here. Illustrative are the many comparisons of inconsistent results observed in the literature, which could benefit from revised interpretations using the found characteristics here. To allow for better comparison and interpretation of results, specific characterizations of stressors and study should be reported.

Limitations and Further Research

There are related issues we have not been able to address in this study that are relevant to examine in further research. First, there are various game and study characteristics that have not been analyzed in our regression models since these had too little variation (see Table A1). In future research, these variables could be included in experimental studies to examine the effect on different physiological stress responses.

Second, three of the characteristics that are significant predictors of physiological stress responses—realism, aversive stimuli, and reliability of effect sizes—are defined using multiple factors. Although this composite definition provides more power to detect moderating effects of game characteristics, it does not provide insight into the moderating effects of the specific factors. In future research, these factors could be examined separately for the potential moderating functions inherent to these factors.

Third, we performed relatively many analyses for the included number of studies. Although the analyses were performed on different physiological variables and different sets of studies, it is important to realize that the used number of analyses gives the performed analyses a more exploratory character.

Fourth, we determined the effect size of a physiological variable based on the measurement that was performed closest to the mean measurement time in the overall study sample. Although this minimizes the effects of unwanted confounding variables, such as habituation or mediating endocrine responses, to be incorporated in the effect size, it is also a limitation in that the effect size is now based on mean measurement values over different time spans poststressor onset.

Fifth, our used method of deriving a p value of .05 for studies explicitly reporting a significant effect, and a $g*$ value of 0 for those reporting no significance, potentially underestimates the calculated effect sizes for the studies for which this method was applied. This is because the used values of .05 for p and 0 for $g*$ are the lower limits of the descriptions given by the authors. This follows that the found effect sizes in this analysis are potentially slightly underestimated.

Conclusion

Digital game stressors can elicit large stress responses and have been used widely in research for several decades (our initial search showed 5,448 articles since 1976). Currently, the importance of the digital game stressor and study characteristics is not

fully recognized. We showed in this article that both game and study characteristics indeed are of importance: variation in single characteristics can alter the effect size by more than one standard deviation and sets of characteristics on average account for 43% of the variation in the stress response in HR and DBP and SBP. These results are consistent and specific, meaning that for different physiological responses different sets of moderating characteristics can be found, and that the findings are consistent across analyses and subgroups. Even more, we showed that, when studies utilize methodology scoring high on the scale introduced in this paper (reliability of effect size), the explained variance increases to 53%. This indicates that this scale can be used to gain more insight in, and control over, the true physiological stress responses and underlying processes. Additionally, we have aimed to map these underlying processes to the different relations between characteristics and physiological responses we

have found, and presented these in one integrated overview. Based on these findings, we propose the notion that (cardiovascular) stress responses are more specific to the type of stressor than previously assumed, and digital game stressors should be viewed as container constructs instead of stressors in themselves. From these insights, two implications may arise. The first is the need for a standardization of digital game stressors. This will aim to overcome the differences in moderating characteristics of digital game stressors and methodology, facilitating the synthesis of research results and the understanding of underlying stress response systems. The second is that researchers designing physiological stress experiments can utilize the findings from this analysis to design experiments and stressors that will provide more consistent and insightful results, and can use these findings for interpreting and comparing results gathered from physiological stress research using digital game stressors.

References

Adams, E. (2010). *Fundamentals of game design 2nd edition*. Berkeley, CA: New Riders.

Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal and prosocila behaviour: A meta-analytic review of the scientific literature. *Psychological Science*, *12*, 353–359.

Arriaga, P., Esteves, F., Carneiro, P., & Monteiro, M. B. (2006). Violent computer games and their effects on state hostility and physiological arousal. *Aggressive Behavior*, *32*, 358–371.

Baldaro, B., Tuozzi, G., Codispoti, M., Montebarocci, O., Barbagli, F., Trombini, E., ... Rossi, N. (2004). Aggressive and non-violent videogames: Short-term psychological and cardiovascular effects on habitual players. *Stress and Health*, 203–208.

Barlett, C., Branch, O., Rodeheffer, C., & Harris, R. (2009). How long do the short-term violent video game effects last? *Aggressive Behavior*, 225–236.

Barlett, C. P., Anderson, C. A., & Swing, E. L. (2009). Video game effects—Confirmed, suspected, and speculative. *Simulation & Gaming*, *40*, 307–433.

Barlett, C. P., Harris, R. J., & Baldassaro, R. (2007). Longer you play, the more hostile you feel: Examination of first person shooter video games and aggression during video game play. *Aggressive Behavior*, *33*, 486–497.

Barlett, C. P., Harris, R. J., & Bruey, C. (2008). The effect of the amount of blood in a violent video game on aggression, hostility, and arousal. *Journal of Experimental Social Psychology*, *44*, 539–546.

Barlett, C. P., & Rodeheffer, C. (2009). Effects of realism on extended violent and nonviolent video game play on aggressive thoughts, feelings, and physiological arousal. *Aggressive Behavior*, *35*, 213–224.

Bernardi, L., Porta, C., Casucci, G., Balsamo, R., Bernardi, N. F., Fogari, R., ... Sleight, P. (2009). Dynamic interactions between musical, cardiovascular, and cerebral rhythms in humans. *Circulation* *119*, 3171–3180.

Berne, R. M., & Levy, M. N. (2001). *Cardiovascular physiology*. Philidelphia, PA: Mosby.

Bradley, M. M., & Lang, P. J. (1999). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings* (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real-life violence. *Journal of Experimental Social Psychology*, *43*, 489–496.

Carroll, D., Turner, J. R., Lee, H. J., & Stephenson, J. (1984). Temporal consistency of individual differences in cardiac response to a video game. *Biological Psychology*, *19*, 81–93.

Codispoti, M., Bradley, M. M., & Lang, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, *38*, 474–478.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Marwah, NJ: Lawrence Erlbaum Associates, Inc.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. Chicago, IL: Rand McNally College Publishing Company.

Dahlquist, L. M., Herbert, L. J., Weiss, K. E., & Jimeno, M. (2010). Virtual-reality distraction and cold-pressor pain tolerance: Does avatar point of view matter? *Cyberpsychology, Behavior and Social Networking*, *13*, 587–591.

DeCoster, J. (2009). Meta-analysis notes. Retrieved from http://www.stat-help.com/notes.html

Dembroski, T. M., MacDougall, J. M., Cardozo, S. R., Ireland, S. K., & Krug-Fite, J. (1985). Selective cardiovascular effects of stress and cigarette smoking in young women. *Health Psychology*, *4*, 153–167.

Dembroski, T. M., MacDougall, J. M., Shields, J. L., Petitto, J., & Kushene, R. (1978). Components of the type A coronary-prone behavior pattern and cardiovascular responses to psychomotor performance challenge. *Journal of Behavioral Medicine*, *1*, 159–176.

Department of Physics, University of South Alabama. *Plot digitizer download*. Retrieved from http://www.southalabama.edu/physics/software/plotdigitizer.htm

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 355–391.

Ditto, B., & Miller, S. B. (1989). Forearm blood flow responses of offspring of hypertensives to an extended stress task. *Hypertension*, 181–187.

ESRB. (2012). *ESRB ratings*. Retrieved from http://www.esrb.org/index-js.jsp

Field, A. (2009). *Discovering statistics using SPSS*. London, England: Sage Publications.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*, 665–694.

Goodie, J. L., & Larkin, K. T. (2001). Changes in hemodynamic response to mental stress with heart rate feedback training. *Applied Psychophysiology and Biofeedback*, *26*, 293–310.

Graham, F. K. (1992). *Attention: The heartbeat, the blink, and the brain. Attention and information processing in infants and adults: Perspectives from human and animal research*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Grossman, P., & Svebak, S. (1987). Respiratory sinus arrhythmia as an index of parasympathetic cardiac control during active coping. *Psychophysiology*, *24*, 228–235.

Hastrup, J. L., & Light, K. C. (1984). Sex differences in cardiovascular stress responses: Modulation as a function of menstrual cycle phases. *Journal of Psychosomatic Research*, *28*, 475–483.

Hebert, S., Beland, R., Dionne-Fournelle, O., Crete, M., & Lupien, S. (2005). Physiological stress response to video-game playing: The contribution of built-in music. *Life Sciences*, *76*, 2371–2380.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*, 107–128.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London, England: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486–504.

Huesmann, R. L., & Bushman, B. J. (2006). Short-term and long-term effects of violent media on aggression in children and adults. *Archives of Pediatrics & Adolescent Medicine*, *160*, 348–352.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. London, England: Sage Publications.

Ironson, G. H., Gellman, M. D., Spitzer, S. B., Llabre, M. M., De Carlo Pasin, R., Weidler, D. J., . . . Schneiderman, N. (1989). Predicting home and work blood pressure measurements from resting baselines and laboratory reactivity in black and white americans. *Psychophysiology*, *26*, 174–184.

Ivory, J. D., & Kalyanaraman, S. (2007). The effects of technological advancement and violent content in video games on players? Feelings of presence, involvement, physiological arousal, and aggression. *Journal of Communication*, *57*, 532–555.

Jones, F., & Bright, J. (2001). *Stress. Myth, theory and research*. Harlow, England: Pearson Education Limited.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The Trier Social Stress Test: A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*, 76–81.

Lawler, K. A., Wilcox, Z. C., & Anderson, S. F. (1995). Gender differences in patterns of dynamic cardiovascular regulation. *Psychosomatic Medicine*, *57*, 357–365.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Technical Report A-8. University of Florida, Gainesville, FL.

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*, 27–50.

Lim, S., & Reeves, B. (2010). Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human Computer Studies*, *1*, 57–68.

Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis*. London, UK: Sage Publications.

Lombard, M., & Ditton, T. B. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*.

Lovallo, W. R. (2004). *Stress and health: Biological and psychological interactions*. Thousand Oaks, CA: Sage Publications.

Lovallo, W. R. (2005). *Stress & health*. Thousand Oaks, CA: Sage Publications.

Markovitz, J. H., Raczynski, J. M., Wallace, D., Chettur, V., & Chesney, M. A. (1998). Cardiovascular reactivity to video game predicts subsequent blood pressure increases in young men: The CARDIA study. *Psychosomatic Medicine*, *60*, 186–191.

Matthews, K. A., & Stoney, C. M. (1988). Influences of sex and age on cardiovascular responses during stress. *Psychosomatic Medicine*, *50*, 46–56.

McFarland, R. A. (1985). Relationship of skin temperature changes to the emotions accompanying music. *Biofeedback and Self-Regulation*, *10*, 255–267.

McKinney, M. E., Miner, M. H., Ruddel, H., McIlvain, H., Witte, H., Buell, J. C., . . . Eliot, R. S. (1985). The standardized mental stress protocol: Test-retest reliability and comparison with ambulatory blood pressure monitoring. *Psychophysiology*, *22*, 453–463.

Miller, S. (1993). Cardiovascular reactivity in anger-defensive individuals: the influence of task demands. *Psychosomatic Medicine*, *55*, 78–85.

Miller, S., & Ditto, B. (1988). Cardiovascular responses to an extended aversive video game task. *Psychophysiology*, *25*, 200–206.

Miller, S. B. (1991). Affective moderators of the cardiovascular response to stress in offspring of hypertensives. *Journal of Psychosomatic Research*, *36*, 149–157.

Miller, S. B. (1994). Parasympathetic nervous system control of heart rate responses to stress in offspring of hypertensives. *Psychophysiology*, *31*, 11–16.

Miller, S. B., & Ditto, B. (1989). Individual differences in heart rate and peripheral vascular responses to an extended aversive task. *Psychophysiology*, *26*, 506–513.

Miller, S. B., & Ditto, B. (1991). Exaggerated sympathetic nervous system response to extended psychological stress in offspring of hypertensives. *Psychophysiology*, *28*, 103–113.

Miller, S. B., Friese, M., & Sita, A. (1995). Parental history of hypertension, sodium loading, and cardiovascular response to stress. *Psychosomatic Medicine*, *57*, 381–389.

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers*, *22*, 336–343.

Newport, D. J., & Nemeroff, C. B. (2002). Stress. In V. S. Ramachandran, (Ed.), *Encyclopedia of the human brain*. (Vol. *4*, pp. 449–462). Amsterdam, The Netherlands: Elsevier.

Nyklicek, I., Thayer, J. F., & van Doorren, L. J. (1997). Cardiorespiratory differentation of musically-induced emotions. *Journal of Psychophysiology*, *11*, 304–321.

PEGI. (2012, 12 12). *PEGI Pan European Game information—Welcome*. Retrieved from http://www.pegi.info/en/

Perkins, K. A., Leonard, E. H., Jennings, R. J., & Stiller, R. (1986). The cardiovascular effects of nicotine during stress. *Psychopharmacology*, *90*, 373–378.

Pollak, M. H. (1994). Heart rate reactivity to laboratory tasks in two daily life settings. *Psychosomatic Medicine*, *56*, 271–276.

Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., & Kivikangas, M. (2006). Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, *15*, 381–392.

Ritterfeld, U., Cody, M., & Vorderer, P. (Eds.). (2009). *Serious games: Mechanisms and effects*. London, England: Routledge.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. London, England: Sage Publications.

Saab, P. G., Tischenkel, N., Spitzer, S. B., Gellman, M. D., Pasin, R. D., & Schneiderman, N. (1991). Race and blood pressure status influences cardiovascular responses to challenge. *Journal of Hypertension*, *9*, 249–258.

Selander, J., Bluhm, G., Theorell, T., Pershagen, G., Babisch, W., Seiffert, I., . . . Järup, L. (2009). Saliva cortisol and exposure to aircraft noise in six European countries. *Environmental Health Perspectives*, *117*, 1713–1717.

Sherry, J. L. (2001). The effects of violent video games on aggression. *Human Communication Research*, *27*, 409–431.

Sims, J., & Carroll, D. (1990). Cardiovascular and metabolic activity at rest and during psychological and physical challenge in normotensives and subjects with mildly elevated blood pressure. *Psychophysiology*, *27*, 149–156.

Sita, A., & Miller, S. B. (1996). Estradiol, progesterone and cardiovascular response to stress. *Psychoneuroendocrinology*, *21*, 339–346.

Skosnik, P. D., Chatterton, R. T., Swisher, T., & Park, S. (2000). Modulation of attentional inhibition by norepinephrine and cortisol after psychological stress. *International Journal of Psychophysiology*, *36*, 59–68.

Slater, M., & Usoh, M. (1993). The influence of a virtual body on presence in immersive virtual environments. *Proceedings of the Third Annual Conference on Virtual Reality*. London, England (pp. 34–42).

Smith, A., Whitney, H., Thomas, M., Perry, K., & Brockman, P. (1997). Effects of caffeine and noise on mood, performance and cardiovascular functioning. *Human Psychopharmacology*, *12*, 27–33.

Smits, P., Temme, L., & Thien, T. (1993). The cardiovascular interaction between caffeine and nicotine in humans. *Clinical Pharmacology & Therapeutics*, *54*, 194–204.

Sokhadze, E. M. (2007). Effects of music on the recovery of autonomic and electrocortical activity after stress induced by aversive visual stimuli. *Applied Psychophysiology and Biofeedback*, *32*, 31–50.

Staude-Muller, F., Bliesener, T., & Luthman, S. (2008). Hostile and hardened? An experimental study on (de)-sensitization to violence and suffering through playing video games. *Swiss Journal of Psychology*, *67*, 41–50.

Stegeren, A. H., Wolf, O. T., & Kindt, M. (2008). Salivary alpha amylase and cortisol responses to different stress tasks: Impact of sex. *International Journal of Psychophysiology*, *69*, 33–40.

Steptoe, A., Melville, D., & Ross, A. (1984). Behavioral response demands, cardiovascular reactivity, and essential hypertension. *Psychosomatic Medicine*, *46*, 33–48.

Stoney, C. M., Matthews, K. A., McDonald, R. H., & Johnson, C. A. (1988). Sex differences in lipid, lipoprotein, cardiovascular, and neuroendocrine responses to acute stress. *Psychophysiology*, *25*, 645–656.

Tafalla, R. J. (2007). Gender differences in cardiovascular reactivity and game performance related to sensory modality in violent game play. *Journal of Applied Social Psychology*, 37, 2008–2023.

Thompson, F. T., & Levine, D. U. (1997). Examples of easily explainable suppressor variables in multiple regression research. *Multiple Linear Regression Viewpoints*, 24, 11.

Tischenkel, N. J., Saab, P. G., Schneiderman, N., Nelesen, R. A., Pasin, R. D., Goldstein, D. A., ... Weidler, D. J. (1989). Cardiovascular and neurohumoral repsonses to behavioral challenge as a function of race and sex. *Health Psychology*, 8, 503–524.

Turner, R. J., & Carroll, D. (1985). Heart rate and oxygen consumptiom during mental arithmetic, a video game and graded exercise: Further evidence of metabolically-exaggerated cardiac adjustments? *Psychophysiology*, 22, 261–267.

van der Vijgh, B. H., Beun, R. J., van Rood, M. M., & Werkhoven, P. (2014). GASICA: Generic automated stress induction and control application design of an application for controlling the stress state. *Frontiers in Neuroscience*, 8, 400. doi:10.3389/fnins.2014.00400

Van Egeren, L. F., Fabrega, H. Jr., & Thornton, D. W. (1983). Electrocardiographic effects of social stress on coronary-prone (type A) individuals. *Psychosomatic Medicine*, 45, 195–203.

Verona, E., Sadeh, N., & Curtin, J. J. (2009). Stress-induced asymmetric frontal brain activity and aggression risk. *Journal of Abnormal Psychology*, 118, 131.

Ward, M. M., Chesney, M. A., Swan, G. E., Black, G. W., Parker, S. D., & Rosenman, R. H. (1986). Cardiovascular responses in Type A and Type B men to a series of stressors. *Journal of Behavioral Medicine*, 9, 43–49.

Whitworth, J. A., Mangos, G. J., & Kel, J. J. (2000). Cushing, cortisol, and cardiovascular disease. *Hypertension*, 36, 912–916.

Wolfson, S., & Case, G. (2000). The effects of sound and colour on responses to a computer game. *Interacting with Computers*, 13, 183–192.

Zachariae, R., Jorgensen, M. M., Bjerring, P., & Svendsen, G. (2000). Autonomic and psychological responses to an acute psychological stressor and relaxation: The influence of hypnotizability and absorption. *International Journal of Clinical and Experimental Hypnosis*, 48, 388–403.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table A1: Overview of potentially relevant game and study characteristics.

Table A2: Overview of construction of main analysis regression models for each physiological variable.

Table A3: Results of tests performed to check assumptions needed to generalize findings of predictor analyses beyond our study.

Table A4: Overview of construction of regression models for subanalysis.