SOMETHING IS ROTTEN IN THE STATE OF AGGRESSION RESEARCH:

NOVEL METHODOLOGICAL AND THEORETICAL APPROACHES

TO RESEARCH ON DIGITAL GAMES AND HUMAN AGGRESSION

Inauguraldissertation
zur
Erlangung des Doktorgrades
der Humanwissenschaftlichen Fakultät
der Universität zu Köln
nach der Promotionsordnung vom 10.05.2010

vorgelegt von
Malte Elson
aus
Köln

9. April 2014

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Juli 2014 angenommen.

I can't believe schools are still teaching kids about the null hypothesis.

I remember reading a big study that conclusively disproved it years ago.

- Randall Munroe, xkcd

Acknowledgments

For always helping me to make the right decisions, I thank Natalie, my (highly) significant other. And I'm in debt to my parents for unconditionally supporting me when I made the wrong ones. I thank Johannes, my partner in crime, for keeping our research on course, and offering adequate distraction from it whenever necessary.

I was truly lucky to have my two supervisors, Prof. Dr. Gary Bente and Prof. Dr. Christopher Ferguson, who allowed me all the freedom in the world in my research adventure, but always offered guidance whenever I wandered off (on side quests). None of this would have been possible, of course, without my boss, Thorsten, who paid not only for my bread and butter, but also for many studies, conference trips, and other events that gave me the opportunity to make new friends abroad (looking at you, Jimmy). And, of course, I want to express my gratitude to Julia, who became my "academic nanny" when legal complications hampered my full adoption.

Although they already know they have my gratitude, I cannot say thank you too often to my co-authors Rohangis, Michael, and Jan (in addition to all the wonderful people mentioned above). A mysterious double-blinded thank you note goes out to the anonymous reviewers that helped advancing our manuscripts until they were publishable. Of course, the work of my student assistants Johannes (another one) and Christoph was invaluable in the process that lead to each paper, as was, of course the time and effort of each of my study participants.

A personal thank you to Dennis, who constantly reminded me why psychology is not one of the natural sciences. To Diana, for being one of the most enjoyable people to hang around with in science since we worked together as student assistants. And of course to my dear colleagues Emese, Elli, Rachel, Florian, Anil, Carla, Ruth, and Jens who have been nothing but good friends since the moment we met. You guys make being at work feel like home, and being at home feel like work.

Table of Contents

Summary

This dissertation offers a comprehensive critique of the current state of research on violent game playing and aggressive outcomes. It discusses twenty-five years of research on violence in digital games and aggression, including empirical evidence, theoretical perspectives, and the heated debates in both the public and academia. The main focus here is on methodological issues limiting the conclusiveness of the research, particularly experiments conducted in psychological laboratories. By suggesting methodological advancements in the study of game violence effects, the thesis wants to offer new perspectives on digital games and aggression to move forward the field and the ideological debates that surround it. The thesis comprises a total of 5 peer-reviewed journal articles (of which 3 are published, one is accepted and in press, and one is under review) that include data from one original study and a secondary analyses of 3 further studies.

The first part of the thesis consists of a detailed review of the current scientific literature on violent game effects with a focus on the theories that have been developed to explain the relationship between the use of digital games and aggression. Important theoretical shortcomings and fallacies of social-cognitive perspectives on how aggression is acquired through violent media contents are identified and discussed.

The second part is a methodological critique of laboratory experiments in research on the effect of violent games. First, common problems and pitfalls in the manipulation of violence as an independent variable and improper control of relevant confounding factors are discussed. The modification of game content ("modding") is suggested as a novel method to meet the requirements of rigorous internal validity and sufficient external validity in psychological laboratory experiments. The advantages of this method are illustrated by the results of an experiment in which it was used. This is followed by an examination of one of the most popular laboratory measures of aggressive behavior (the Competitive Reaction

Time Task), providing evidence from three studies that the unstandardized use in the scholarly literature poses a threat to its interpretability and generalizability.

The dissertation concludes with an analysis of the scientific discourse on the game violence-aggression link, and the ways in which it is shaped by ideological convictions that affect both the theoretical assumptions and the methodological procedures. This duality of ideologies present in theory and methods constitutes a threat to violent game effects research, as it causes the field to stagnate. It is argued that this stagnancy can only be resolved through methodological rigor that will, ultimately, advance inadequate theories of media effects.

Introduction

The debate about harmful effects of popular media began long before the television or the personal computer entered our everyday life. During the first half of the 20th century it were mostly the radio and comic books that came under criticism (Ferguson, 2013a), but already in the late 18th century the so-called "reading mania", particularly regarding the excessive consumption of fiction novels by females, caused a heated dispute. Now, as then, there are two sides arguing whether and how morally objectionable media content (such as displays of violence) can affect its users. While the link between exposure to violent media and aggression has been debated in academia for decades (J. Anderson, 2008), the rise of the medium of digital games as a means for media users to *cause* (and enjoy) on-screen violence with a simple press of a button sparked a new controversy in science and the public, that has been ongoing for more than twenty-five years now.

Researchers on both sides of this debate (sometimes dubbed "believers" and "skeptics") have a great interest in understanding the antecedents and underlying mechanisms of aggression, as well as in reducing violence in society. Many scholars are convinced that a strict regulation of violence in games would lead directly to a reduction of problematic behaviors and crime rates, and that those opposing regulation induce a societal risk (e.g., Huesmann & Skoric, 2003). Others worry that the debate on media violence could distract from societal issues they consider more relevant to the etiology of aggression, such as poverty or inequality, and ultimately cause harm as these factors get ignored (Ferguson, 2013b). This tension between groups of scholars who are, naturally, convinced of the validity of their scientific opinion, but who also consider the behavior of the 'other side' to be harmful or dangerous, is a breeding ground for a heated debate often torn by ideological convictions (Grimes, J. Anderson, & Bergen, 2008). Yet both sides agree that science can only provide an answer to the question whether violence in digital games can cause aggression on two conditions:

1. Theories and models must be able to accurately describe the effect mechanisms of violent games and offer predictions from which testable, falsifiable hypotheses can be derived.
2. Empirical operationalizations of these mechanisms and the measurement of aggression need to be objective, reliable, and valid, so that they can be properly interpreted and generalized.

One major reason for disagreement among scholars, however, is a dissent whether (or to what extent) these two conditions have been met.

The belief that exposure to violence in games increases aggression has been closely tied to a social-cognitive perspective on media effects and aggression, specifically Bandura's (1978) observational social learning theory. The underlying assumption is, in essence, that media characters function as models for behavior and that humans learn through the observation of game avatars just like they learn through the observation of others in their physical environment. Thus, observing on-screen violent behaviors being rewarded or punished would respectively increase or decrease aggressive tendencies in game players. Given that the playing of many games, e.g. the notorious first-person shooters, is somehow tied to the killing of opponent avatars, one might argue that, theoretically, the conditions for this mechanism are met. A popular formula for testing these assumptions empirically in laboratory experiments is to have study participants play one of two games (a violent and a nonviolent one), after which they partake in a laboratory procedure intended to measure aggressive behaviors. This “boilerplate” for experiments has been replicated dozens, if not hundreds of times, and makes up large parts of the foundation of empirical evidence on digital game violence effects (and media violence effects in general).

However, some media effects researchers have expressed doubts regarding the usefulness of conventional social-cognitive theories in predicting media violence effects, and criticized them for offering too simplistic views on the etiology of human aggression (Ferguson & Dyck, 2012). There are also profound concerns whether common empirical approaches to studying these effects in psychological laboratories, particularly regarding the methodology of measuring aggression, allow drawing meaningful conclusions about media violence (Ritter & Eslea, 2005; Tedeschi & Quigley, 1996).

This dissertation critically examines four theoretical and methodological aspects in game violence effects research: (1) extant models and perspectives predicting relationships between violent games and aggression, (2) the manipulation of independent variables and control of confounds in experiments on game effects, (3) the measurement of aggression in psychological laboratories, and (4) ideological biases that shape both the research as well as the academic debate that surrounds it.

The first paper (Elson & Ferguson, 2014a) included in this dissertation was published in *European Psychologist* and presents an overview of the theoretical and empirical literature on the effects of displayed violence in digital games on aggressive cognitions, emotions, and behaviors. It focuses on theoretical shortcomings of conventional, social-cognitive views on aggression and media violence effects, and discusses relevant gaps in the empirical scholarship. This review offers some new perspectives on potential mechanisms of media effects that could guide future empirical research in this area. Four scholars with opposing or diverging views were invited to comment on this article (Bushman & Huesmann, 2014; Krahé, 2014; Warburton, 2014) and offer a rebuttal to the issues that were raised.

The second paper (Elson & Ferguson, 2014b) is a response to these comments and was, hence, also published in *European Psychologist*. Due to the nature of the three comments, this article reiterates some of the theoretical concerns in greater detail, but largely constitutes a discussion of the scientific discourse around violent digital games research rather than the science itself. It offers explanations why there is such a heated debate around violent game effects in the scientific literature, how this might pose a threat to the credibility of media effects research in general and its ability to inform society about a topic of public interest in particular.

The third article (Elson & Quandt, in press), to be published in *Psychology of Popular Media Culture*, constitutes a transition from theoretical considerations to novel methodological approaches to studying game effects. The article presents a rationale why using multiple games (violent and nonviolent ones) to create different conditions in laboratory experiments in order to study the effects of one specific variable (violence) violates fundamental assumptions of experiments as a scientific method. Put briefly, by using different games to manipulate the target variable of violent content, one is likely to accidentally manipulate other variables in this process that could easily conflate or confound any findings on dependent variables, if not properly controlled for. Next to a thorough examination of this problem and its prevalence in experiments with games, the article offers game modifications ("modding") as a viable solution readily available to all game researchers.

The fourth article (Kneer, Elson, & Knapp, under review) corroborates these methodological considerations with empirical data from a 2x2 experiment in which the violent content and difficulty of one game were manipulated (while holding all other game characteristics constant) instead of using different games for each condition. The data confirm the assumption that game difficulty is a key variable when studying the effects of game violence, particularly regarding emotional responses, and that modifying games constitutes a useful approach to studying the effects of individual game variables.

The fifth paper (Elson, Mohseni, Breuer, Scharkow, & Quandt, 2014), published in *Psychological Assessment*, concludes this dissertation by raising methodological concerns regarding the measurement of aggression in laboratory experiments. The paper discusses psychometric properties of one of the most widely used paradigms to measure aggression, the Competitive Reaction Time Task, and focuses specifically on objectivity and standardization issues. These concerns are confirmed by data from three studies that show that analyzing the same data with different variants of this test leads to large differences in significance levels, effect sizes, and even the direction of effects. Implications for the empirical literature on violent digital games and aggression research in general are discussed and practical suggestions on how this test should be used in order to arrive at more objective results are provided.

Learning to be Aggressive from Violent Games

Given the large popularity of digital games, the implications of their presumed effects on aggression would be unsettling. For example, at least one quarter of the German population plays digital games (Quandt, Breuer, Festl, & Scharkow, 2013). If the use of digital games did, indeed, have an effect on aggression, this would imply a substantial societal problem. But what exactly are the purported psychological mechanisms that would make people more aggressive from their exposure to violent digital games? Historically, many researchers have defaulted to a socio-cognitive perspective on the etiology of aggression in general and specific to the role of media violence. The following sections present an overview of proposed socio-cognitive mechanisms and their shortcomings, while arguing for the adoption of theories focused on biological determinants and influences from the social environment (such as family and peer groups).

Definitions of Aggression and Media Violence

The disagreement whether violent games cause aggression begins in basic scientific questions, such as the definitions of violence and aggression. The field of media violence effects research started out with operational definitions of aggression focused purely on the outcome of behaviors. Buss (1961), for example, defined aggression as one organism presenting painful stimulation to another organism. While this behavioristic approach to aggression had great merits for researchers in practice, it was eventually considered insufficient as it could not distinguish between accidental and intentional behaviors causing harm (which of course can be crucial, for example, in court decisions on crimes). A commonly used definition of aggression is the one by Baron and Richardson (1994), who defined aggression as any behavior that is intended to cause harm to another person who intends to avoid this harm. However, Grimes et al. (2008) cautioned against defining aggressive behaviors through preceding intentions, as it might entice psychologists to measure cognitive processes underlying aggression rather than whether any harm or injury has actually been inflicted. Grimes et al. (2008) criticize that, as violent behaviors are simply considered extreme forms of aggression (C. Anderson & Bushman, 2002), measuring these cognitive variables is, accordingly, also a sufficient proxy for violent behaviors, or even violent crimes. To address this problem, van der Dennen (1980) suggested to separate aggression and violence as completely distinct categories. According to this definition, as long as a drive, impulse, or desire to inflict pain is operative, a behavior should be considered aggressive (motivational component). Violence, on the other hand, describes a category of behaviors involving harm, elimination, or destruction, which can be direct or indirect, and physical or mental (behavioral component). As such, there are four types of behaviors: Aggressive violent behaviors (e.g., a crime of passion), aggressive nonviolent behaviors (e.g., gossip), nonaggressive violent behaviors (e.g., executions), and, of course, non-aggressive non-violent behaviors (the residual category).

Defining media violence or determining how violent one game is compared to another seems even more intricate, and varies substantially. The reason why psychological definitions of aggression or violent behaviors, such as the ones discussed above, might not be applicable to games is that a) since all game violence is virtual, and not physical, it is questionable whether any harm is actually being inflicted, and b) the narrative intentions of avatars might not be congruent with the intentions of players. Usually, empirical publications do not provide a definition at all since the difference in violent contents between games selected for experiments often has a high “face validity” (e.g., when one involves a considerable amount of combat, and the other is an abstract puzzle game). Of course, this theoretical gap becomes apparent when the public turns to psychologists and asks to apply their research to practical decisions. For example, a proper definition of media violence becomes necessary when lawmakers or judges decide whether specific violent contents are harmful enough to warrant censorship or legislation that limits access to games

(Brown v. Entertainment Merchants Association, 2011), or simply when concerned parents wonder which games are and which are not suitable for their children. The increased graphicness of violence through technological advancements has alerted researchers to study potential increases in their effects (e.g., Ivory & Kalyanaraman, 2007), yet in most cases it does not seem to be a relevant factor when it comes to defining what makes a game violent. At least historically, researchers seemed more concerned with whether any violence is being rewarded in games rather than the magnitude of violent content itself, and considered seemingly innocuous titles like *Super Mario* violent as well (C. Anderson & Dill, 2000). According to Ferguson (2014), however, this rather vague perspective could render the category "violent games" useless. Recently, there have been more sophisticated attempts, and scholars suggested describing game violence through multiple technological and narrative components, such as graphicness, realism, and justification (Tamborini, Weber, Bowman, Eden, & Skalski, 2013).

The General Aggression Model

In 2002, C. Anderson and Bushman published a revised version of the General Aggression Model (GAM), a synthesis of several social-cognitive and neoassociative theories, that has become the default model for many game violence researchers, particularly for those who believe games to be a strong cause for aggressiveness. The GAM is strongly rooted in Bandura's (1978) social learning theory (SLT) of aggression, which predicts that aggressive behaviors can be reinforced either through direct experience or vicarious observation of aggressive acts being rewarded. The greater the rewards, the greater the reinforcement and, consequently, the likelier the chance of imitating what has been observed. According to SLT, a repeated experience or observation of aggressive behavior being rewarded results not only in a higher frequency of aggressiveness as the reward expectation increases, but also alters concepts regarding the appropriateness of aggressive behaviors in a wide range of situations. Thus, these rewarded models cause a greater preference for aggression as instrumental to reaching goals, shape aggressiveness as a general response class, and ultimately consolidate it as a social norm.

Through repeated exposure to aggression, the GAM also predicts changes in knowledge structures, such a perceptual and behavioral schemata. As such, aggressive behaviors are also accompanied by an increasingly hostile perception of the world and the presumed intents of other persons. Whether a person responds aggressively or not to a particular event is determined in the GAM's tripartite process model. Situational characteristics (e.g., aggressive cues) and personality variables (e.g., traits, learned scripts) are located on the input side. The interpretation of the environmental input depends on internal states of cognitions (e.g., hostile thoughts), affect (e.g., mood), and psychophysiological arousal. This can be a relatively effortless, impulsive, and automatic process. However, when the immediate appraisal is not satisfactory and resources are not limited (usually time and capacity), any given information can be re-evaluated numerous times. Either way, this immediate or thorough evaluation determines a behavioral response as the outcome. The response to this outcome, again, becomes part of the information for the next episode. Ultimately, repeated episodes of actions and reactions result in more permanent perceptual, attitudinal, and behavioral patterns. And, in accordance with SLT, this mechanism also works when such action-reaction-chains are observed in the behavior of others.

Theoretical Shortcomings of the GAM

The GAM does not differentiate sufficiently between observations in physical and digital environments. The way violence is being rewarded in digital games, e.g. as a necessary condition to win the game, in-game benefits (e.g., better equipment), or through scores on leaderboards, is considered to be sufficient as a reinforcement of aggressive behaviors for game players. Accordingly, repeated exposure to games with such contents would be considered a risk factor in the etiology of aggression. The first publication included in this thesis (Elson & Ferguson, 2014a) discusses at least 5 major weaknesses of the GAM perspective on the development of aggression in general and its specific assumptions about effects of game violence.

1. While offering simple and testable predictions about the antecedents and consequences of human aggression, the GAM is heavily focused on cognitive scripts and does little to elucidate motivational and personological variables that may influence aggressive behaviors. Within the GAM's line of argument, personality characteristics and motivations for behaviors are, in essence, nothing more than strongly and repeatedly reinforced cognitive scripts, thereby rendering it a "tabula rasa" theory (Pinker, 2002).
2. Particularly biological and genetic factors are neglected despite their importance in predicting aggression and even crime in individuals (Ferguson, Ivory, & Beaver, 2013). It remains especially unclear how they interact with supposedly acquired aggressive scripts. There is also a lack of specific variables explaining individual susceptibility or immunity to potential effects of violent games.
3. Despite its popularity in psychological media effects research, the GAM is not actually used by clinicians or other professionals in the field dealing with pathological forms of aggression (Ferguson & Dyck, 2012). Neither are there clinical diagnostic instruments based on the GAM, nor is the GAM being used to inform programs aimed at reducing pathological aggression, as opposed to biopsychosocial models of aggression that dominate clinical psychology.
4. The GAM does little to account for competing schemata and scripts. Even assuming that violent games are able to model aggression in their players, these models will sooner or later either be contradicted by punishments for aggressive behaviors, or contested by rewards for prosocial models, for example through

parents or peers. The GAM does neither predict the outcome of competing models, nor which kind of rewards for aggressive scripts might supersede nonaggressive scripts (and vice versa).

5. The GAM equates the effects of observations of rewarded aggression in virtual and physical environments, and predicts that it makes no difference for the underlying mechanisms to work whether the observed violence is fictional or real.

The fifth point deserves some further attention, since Bushman and Huesmann (2014) responded to it specifically by arguing that the assumed equality of learning opportunities constitutes a theoretical advantage of the GAM, not a limitation. Bushman and Huesman explicitly ask for a theory that would explain how viewing violence mass media could be different from, for example, observing violence in war-torn countries (Boxer et al., 2013). Therefore, the second publication (Elson & Ferguson, 2014b) expatiates on this argument further. Briefly, assuming that these two experiences could be similar is faulty on three grounds:

5a. The experience or observation of fictional acts of violence (e.g., a knight killing a dragon) is not similar in its qualities to the experience and observation of real violence (e.g., a news report on an ongoing war), even when both are presented on screen. This is corroborated by evidence showing that children at the age of five (or younger) are already able to distinguish between real and fictional television (Wright, Huston, Reitz, & Piemyat, 1994).

5b. Observing violence in digital games does not have similar psychological effects as observing proximal acts of violence (e.g., in the family), even when both acts would be similar. This is substantiated by a large body of research findings (e.g., Ferguson, San Miguel, & Hartley, 2009) as well as decades of clinical practice and psychological assessments of children witnessing domestic violence in their families (Levendosky, Huth-Bocks, Semel, & Shapiro, 2002).

5c. There is no theoretical explanation how observations of violence in virtual worlds (e.g., soldiers fighting) generalize to actual behaviors that are completely different in the real world (e.g, domestic abuse). While one could make a strong case for how digital games teach that violence is a promising and successful measure in other similar games, the transfer from behaviors in digital environments to other behaviors in physical environments (that are different in many aspects from the virtual environments) is at least not a natural given.

Finally, even ignoring all the issues raised above and assuming that repeated exposure to violence games would incrementally make players more aggressive, it must be taken into account how others would react to these changes in behaviors. A progressive increase in aggressive behaviors would usually get punished by peers or the family, thus decreasing them through undesired consequences. If, however, aggressive or antisocial behaviors are tolerated or even rewarded, does not the real issue lie within an unhealthy environment that fosters aggressiveness rather than peacefulness?

Violent Digital Games: Manipulation and Control of a Multifaceted Stimulus

Taken together, the first two publications (Elson & Ferguson, 2014a, 2014b) offer an overview of the results obtained in laboratories and the field with a strong focus on methodological rigor, and integrate different perspectives and interpretations to explain their relevance to the understanding of media effects. The main body of psychological research on the effects of digital games consists of laboratory experiments. Many of these studies share a certain design: Study participants (typically college students, most of them psychology or communication majors) either play a violent game (mostly a first-person shooter) or a nonviolent game. Psychophysiological arousal (heart rate, skin conductance level) is sometimes measured during, or before and after play. After playing one of the games, participants are subjected to a test or fill out a questionnaire to assess aggressive cognitions, emotions, or behaviors, which are then compared for the two groups. Any observed differences between those groups are then usually explained with the manipulation of violent content. An example for this kind of research design can be found in the study by K. Williams (2009) in which participants played either *Mortal Kombat: Deception* (Midway, 2004) or *Dance Dance Revolution Max 2* (Konami, 2003). *Mortal Kombat* is a fighting game in which players control a character engaged in close combat with an opponent. One match usually involves several rounds of fighting in an arena. *Dance Dance Revolution*, on the other hand, is a rhythm game in which players typically have to mimic dancing instructions to pop songs on special dance mats that serve as input devices. In this study, participants completed a hostility scale after playing and the results show that those who played *Mortal Kombat* reported significantly stronger feelings of hostility. K. Williams (2009) concludes that "[t]his supports past evidence that exposure to violent video games, when compared to nonviolent video games, results in aggressive affect" (p. 303).

However, this way of manipulating game contents as independent variables could violate fundamental assumptions of experiments as a scientific method. In his classic *Experimental Psychology*, Woodworth (1938) describes the defining elements of the scientific method of experiments in psychological science as we still know it today: For a study to qualify as an experiment, the researcher "holds all the conditions constant except for one factor which is his 'experimental factor' or his 'independent variable.' The observed effect is the 'dependent variable' which in a psychological experiment is some characteristic of behavior or reported experience'' (p. 2). And while this assertion is being taught in any ordinary introductory psychology class, it has serious consequences for research on and with digital games.

Arguably, it is very convenient and certainly bears convincing face validity to divide games into two groups according to the current variable of interest (e.g., violent and nonviolent games). Yet with a complex stimulus like games it should be considered that violence is unlikely to be the only difference between two games that have been selected for research purposes. Any of those additional differences constitutes a potentially confounding factor that might bias results if not controlled for. This problem is particularly intricate as there are common cooccurrences of themes, contents, and mechanisms in certain game genres (Apperley, 2006) that could lead to a systematic conflation in larger bodies of research. Although not a genre in itself, it is certainly possible that violent games commonly share some other characteristics that could be relevant for aggression research.

Adachi and Willoughby (2011b) argue that when investigating effects on aggressiveness, scholars should consider the *difficulty*, *pace of action*, and *competitiveness* of a game as possibly relevant variables besides violent content. A genre that most often features displays of violence is the first-person shooter. These games are usually also fast-paced, likely to be played competitively against other human players, and highly demanding in terms of perception and motor abilities – not to mention that first-person shooters are always played from the first-person perspective. By contrast, the puzzle games (e.g., Tetris) that are popular stimuli for "control groups" are typically nonviolent, but also rather slow-paced, usually played alone, and require cognitive efforts, such as problem-solving abilities and mental rotation. So when observing differences in measurements between those groups after playing, does it mean that one particular game characteristic, such as violence, affected human behavior? This example illustrates that studies attributing changes or group differences in aggression to violent contents specifically might be severely confounded by other contents that were not properly controlled for, or even accidentally manipulated by using different games that varied on multiple dimensions.

The third publication (Elson & Quandt, in press) provides a detailed discussion on the problem of stimulus control in research with digital games. First, advantages and disadvantages of previous approaches to this problem are considered (e.g., using Likert scales to rate and control for relevant third variables), followed by the introduction of game modifications (or "mods") as a viable alternative to manipulate independent variables and control confounding factors. Of the different types of mods Scacchi (2010) identifies, the most relevant to psychological researchers are so-called *partial conversions,* which are smaller alterations or additions to an existing commercial game. These mods range from relatively small and cosmetic additions, such as new textures for existing objects and clothes for characters, to entirely new environments the game can be played in. While mods are usually being created for entertainment purposes, they could arguably be used as powerful manipulations of independent variables, while at the same no other aspects of a game would be changed and thus exerting meticulous control over confounding variables. To assess the relevance of modding for digital games effects research, Elson and Quandt (in press) conducted a three-step systematic literature review with the aim to estimate to what extent researchers already used modding techniques, but also which proportion of the experimental work on games could potentially benefit from it.

First, several academic databases were searched for all peer-reviewed entries using the following three terms in the field *All Text* (TX): *modding*; *video game** and *mods*; *video game** and *modif**. This resulted in a total of 52 publications that either dealt with the topic of modding specifically or utilized modding techniques for research purposes. Naturally, this type of literature search was unable to retrieve publications in which scholars make use of modding techniques without referring to them as such or using the terminology more common among game designers than social scientists. Therefore, the literature review was extended in a second step. The last ten volumes (total number of articles N = 4,160) of the journals *Communication Research* (Sage), *Computers in Human Behavior* (Elsevier), *Cyberpsychology, Behavior, and Social Networking* (Mary Ann Liebert), *Human Communication Research* (Wiley), *Journal of Communication* (Wiley), and *Media Psychology* (Taylor & Francis) were searched for games-related articles to which modding could, in theory, be applied as a means of stimulus creation, manipulation, or control. Of the n = 145 studies that employed digital games as stimuli, 26 (18%) used materials that were modded by manipulating contents so that the games would be more suited for their research questions (e.g., varying contents to create conditions, or removing unwanted contents to exert greater stimulus control). In 42 studies (29%) at least one independent variable was manipulated by using two or more completely different commercial off-the-shelf titles (potentially diminishing internal validity), and for 28 studies (19%) entirely new games were created as stimulus materials (potentially diminishing external validity). At least in these cases, modding one game instead, or using different playing modes of the same game (as suggested by McMahan, Ragan, Leal, Beaton, & Bowman, 2011) might have been viable alternatives.

Psychologists interested in game violence effects have benefited from utilizing mods in laboratory studies in the past. The earliest example is identified is the study by Staude-Müller, Bliesener, and Luthman (2008), whose participants either played a conventional first-person shooter (FPS), or a mod in which avatars are being frozen instead of killed. Of course, this sophisticated mod might not fully solve problems of stimulus control, as it could be questioned whether freezing someone should be considered truly “nonviolent” (after all, like dying, freezing is not a very desirable experience). However, it still constitutes a highly functional approximation of a clear *relative* difference in degrees of violence between conditions. Any outcome variable that differs between the “kill” and the “freeze” version could be attributed to the degrees of manipulation, even when the latter version does not remove *all* violence from the former. Following the suggestions by Adachi and Willoughby (2011b), Elson, Breuer, Van Looy, Kneer, and Quandt (2013) studied the isolated and interaction effects of violence and pace of action in digital games on cardiovascular responses and aggressive behavior, They assigned their participants to play one of four versions of a FPS: normal-paced (default speed level) vs. fast-paced (speed level at 140%), violent (wielding a grenade launcher) vs. nonviolent (wielding a toy nerf gun). Hartmann, Toz, and Brandon (2010) created two mods to assess the effects of (un)justified violence on feelings of guilt: Their participants were either playing UN soldiers attempting to shut down a torture camp, or paramilitary forces defending the camp and continuing the cruelty.

In order to offer researchers a good rationale where to start when considering modding for an upcoming experiment, the third publication (Elson & Quandt, in press) also includes a brief overview of the modding tools currently available. While it cannot replace the study of elaborate tutorials, it aims to provide a rough idea of which tools might and might not suit researchers’ needs.

Game Difficulty as a Relevant Confound in Game Violence Research

Further corroborating the theoretical considerations by Adachi and Willoughby's (2011b) about the relevance of other game characteristics, the fourth publication (Kneer et al., under review) presents an experiment examining the effects of violence and difficulty in digital games. Difficulty is a particularly important variable as an unsatisfactory in-game performance might be frustrating to players. In line with the classic frustration-aggression hypothesis (Berkowitz, 1989; Dollard, Miller, Doob, Mowrer, & Sears, 1939), frustration resulting from a mismatch between game difficulty and player skills could lead to increases in aggression. Conversely, a game being too easy might be quite boring to players, particularly highly skilled ones. Past studies, however, have rarely controlled for game difficulty. When using different games to manipulate violent contents this could present a considerable problem as these games might also differ on their difficulty levels. In these cases, it would be quite problematic to determine whether increases in aggression can be traced back to the level of displayed violence or occur as a negative response to unattainable in-game challenges.

In the study by Kneer et al. (under review), $N = 90$ participants played the first-person shooter *Team Fortress 2* (Valve, 2007) in which two teams both try to capture and hold a control point while preventing the other team from doing so (in this study, teammates and the opponent team were controlled by the computer). Participants were assigned to one of four conditions with either high or low difficulty settings and a high or low amount of violent content. These conditions were created with in-game options and through publicly available modding tools. In the high violent conditions, the player and all bots wielded flamethrowers, and the portrayed deaths of characters in the game were rather bloody and graphic. In the low violent conditions, everyone was equipped with a 'rainbowblower' that blasted rainbows instead of fire while playing bubbling sounds, and instead of dying, this weapon incapacitated characters by making them drop to the ground convulsing with laughter. Difficulty was manipulated by altering the weapon's damage output, the player's resistance to enemy damage, and the speed at which the control point could be captured. Thus, instead of four different games, participants played one of four versions of the same game only differing with regard to the independent variables while holding all other variables constant. The dependent measures were psychophysiological arousal (interbeat intervals and electrodermal activity measured continuously during play), aggression-related associations (measured with a lexical-decision task after play), aggressive behaviors (measured with the standardized Competitive Reaction Time Task after play), as well as positive and negative emotions (measured with Renaud and Unz's [2006] affect scale).

Results show that there was no influence of violent content on psychophysiological arousal, aggression-related associations and aggressive behavior, or positive and negative affect. Difficulty did not have any appreciable effect on psychophysiological arousal, aggressive behavior, and positive or negative affect. However, a higher difficulty was significantly associated with higher response latencies for aggressive words in the lexical decision task. Thus, a higher difficulty inhibited aggression-related associations. This difference was not significant for neutral words but the trend was similar, showing higher response latencies when the difficulty was increased. The reason for this finding might be that a higher difficulty of a game leads to exhaustion, resulting in slower responses in the lexical decision task in general. The interaction of violent content and difficulty did not produce any significant changes in any of the dependent variables. However, the results provide strong evidence that in-game performance (measured through the total number of opponents killed by the participant) predicts both positive and negative affect after play.

This study adds to the emerging literature on game characteristics particularly relevant to violent game effects research, such as pace of action (Elson et al., 2013), competitiveness (Adachi & Willoughby, 2011b), or technological advancement (Ivory & Kalyanaraman, 2007). While this study provides no evidence that game difficulty confounds measures of aggressive behaviors, its results suggest that difficulty and in-game performance should be taken into account when studying cognitive and affective processes during and after the exposure to violent game contents. This is further supported by studies investigating other factors than difficulty that might elicit frustration in game players. For example, the studies by Breuer, Scharkow, and Quandt (2013) and Elson, Breuer, Scharkow, and Quandt (2014) show that, next to the outcome of the game (winning or losing), the behavior of others (e.g., their playing abilities, or their friendliness) can significantly frustrate players, which, in turn, predicts aggressive and cooperative behaviors towards their coplayers. These are, of course, variables related to the playing situation rather than the game itself. Finally, the study by Kneer et al. (under review) can be considered an example of how using options provided by games and modding tools can help psychologists to carefully design experiments that meet the requirements of clean variable manipulation and rigorous variable control.

Measuring Aggression in Laboratories: A Cautionary Tale

After presenting approaches to the precise manipulation of independent variables in violent game effects research, the following section is concerned with methodological concerns regarding the dependent measures, i.e. aggression. A large number of studies investigated the facilitation of aggressive cognitions (e.g., thoughts) through violent digital game playing. Aggressive cognitions themselves, or even simple aggressive thoughts are, however, quite difficult to assess as they cannot be observed directly and would need to be verbalized – or expressed in a different manner – by study participants. Instead, psychologists usually measure superficial correlates of aggressive thoughts, such as automatic semantic activations or the accessibility of words and concepts related to aggression. Popular measures of these associations are, for example, lexical decision tasks (see above) or the word stem completion task, in which participants are presented with a series of ambiguous items that can make more than one word by filling in the respective letters. Depending on which letters are inserted, the meaning of the word can either be related to aggression or to something else (e.g., "k i _ _" having the two possible completions "kill" and "kiss"). The underlying idea is, in essence, that a higher number of aggression-related completions indicates a greater accessibility of aggressive cognitions.

Arguably, these types of measures severely limit the real-world relevance of the results, as they cannot be generalized to actual aggressive thoughts, let alone aggressive behavioral tendencies. While these measures might be helpful when investigating which concepts (including aggression-related ones) are being primed by specific types of games, and which concepts might be suppressed (e.g., Kneer, Glock, Beskes, & Bente, 2012), they do not allow inferring any intent to commit aggression or violent crimes. Unsurprisingly, the majority of studies using these and similar measures do find that games with violent contents increase the accessibility of aggression-related concepts (Barlett, Branch, Rodeheffer, & Harris, 2009; Sestir & Bartholow, 2010) compared to nonviolent games. To what extent these studies contribute to the understanding of violent game effects, however, remains debatable. They certainly do no warrant an alarmist warning of hazardous effects of digital games on the way we think, or even react to stimuli from our environment.

The Unstandardized Use of the Competitive Reaction Time Task

Most of the discussion about the potential harm of violent games within the scientific community, news media, and the general public has focused on the issue of whether violent digital game exposure results in aggressive or violent actions. However, this has been a difficult question to answer. Legal and ethical restrictions make measuring aggressive behavior in a laboratory a difficult enterprise. As can be imagined, it is generally not possible to create a scenario in which individuals will attack each other in the laboratory environment. Unfortunately, this means that most experiments must rely on instruments that do not measure aggression or violence directly, but vaguely approximate it in some way. Notable examples of these measures are the amount of hot sauce used by the participant to spice bowl of chili for someone else (Lieberman, Solomon, Greenberg, & McGregor, 1999), the number of needles used to pierce a voodoo doll (DeWall et al., 2013), or the accuracy of darts thrown at pictures of human faces (Mussweiler & Förster, 2000). An instrument used in many experimental studies (not only in media violence effects research) is the Competitive Reaction Time Task (CRTT). In the original version of the CRTT by Taylor (1967), the Taylor Aggression Paradigm, participants were led to believe that they would be playing 25 consecutive rounds of a reaction time game against another participant in which the winner of a round would punish the loser with an electric shock. Participants who lost a round would receive shocks of varying intensity and when participants won a round they could adjust the shock levels for their alleged opponents. The intensity level of the shock was used as the measure for aggressiveness. Recent adaptations of the CRTT allow participants to set the intensity (usually volume and/or duration) of a noise blast instead of an electric shock, as they are easier to use and bring up fewer ethical issues (Ferguson & Rueda, 2009). As there is no real opponent, the sequence of wins and losses, as well as the settings "chosen" by the opponent, are typically randomized and preset. Generally, louder and longer noise blasts are considered indicators of higher levels of aggressiveness.

However, the CRTT has been used in many different versions in the past. Inconsistencies are found in the procedure of the CRTT, as well as in the ways in which the CRTT data are analyzed by different (and sometimes even the same) authors. While the procedural aspects refer to the setup of the test, i.e., how the raw data are generated, the statistical differences refer to how the data are analyzed. Of course, the procedural decisions also affect the options for statistical analyses. At least 13 different variants to calculate a score for aggressive behavior can be found in the literature: Multiplication of each trial's volume and duration (Bartholow, Sestir, & Davis, 2005), volume and square root of duration (Carnagey & C. Anderson, 2005), or volume and log-transformed duration (Lindsay & C. Anderson, 2000); standardized and summed volume and duration (Bartholow, Bushman, & Sestir, 2006); separate average volume and log-transformed duration settings for each outcome (wins and losses) (C. Anderson & Dill, 2000); average volume, not allowing any duration settings at all (Sestir & Bartholow, 2010); sum of high volume settings, i.e. 8 to 10 on a scale from 1 to 10 (C. Anderson & Carnagey, 2009); separate volume and duration setting of only the first trials (Bushman & Baumeister, 1998); the setting of the first trial and the means of trials 2-9, 10-17, and 18-25 (C. Anderson et al., 2004); volume and duration in a two-phase version of 25 trials each, in which the participant can retaliate in the second phase for the punishment received during the first (Bartholow & C. Anderson, 2002).

From a methodological point of view, inconsistent procedures and analyses are highly problematic because they infringe upon the objectivity criterion of psychological test theory (Kaplan & Saccuzzo, 2009). Simmons, Nelson, and Simonsohn (2011) pointed out that flexibility in data collection, analysis, and reporting in psychological research dramatically increases actual rates of false-positive findings. Moreover, if there is no standardized procedure for a test and no standardized way to process the raw data into a meaningful score, the question remains whether the unstandardized value really approximates the true value of the construct. Aggression scores that are calculated with different procedural versions of the same test become very difficult to compare. Under the assumption that all these different procedures and analyses are equally capable of measuring the construct of aggressiveness, it is unclear why so many versions exist. Without a doubt, theory-driven modifications of a method such as the CRTT, with the aim of answering specific research questions, can contribute to the understanding of psychological processes and extend the area in which a certain test can be applied. However, many authors do not explain in detail why they decided on a specific test procedure or on the aggression score they calculated from the raw data. In many cases, it is not clear why a particular score should be more suitable than others to address the respective research questions. Sometimes, the decision to focus on one of many possible scores seems to have been made post hoc, not prior to data collection.

While there have been several studies that examined at the validity of the test (Ferguson, Smith, Miller-Stratton, Fritz, & Heinrich, 2008; Suris et al., 2004; Tedeschi & Quigley, 1996), until now, there has been no study that addresses the aforementioned objectivity issues by systematically comparing the different analysis procedures for the CRTT. The fifth publication (Elson, Mohseni, et al., 2014) presents data from three studies that were conducted to investigate the effects of digital games on aggressive behavior (measured with the CRTT). All analysis procedures that could be identified in the literature were applied to the three datasets with the aim to investigate whether there would be any variability of results when using different CRTT scores within each study, and whether this variability could be replicated across studies. The analyses showed that there was a considerable range of significance levels (from p = .070 to .934 in study 1; p = < .001 to .959

in study 2; p = .096 to .212 in study 3) and effect sizes (from ω = .0 to .10 in study 1; ω = .0 to .39 in study 2; ω = .09 to .20 in study 3). Thus, it seems that the calculation of different aggression scores can lead to results that are substantially different from each other; in one case, even diametrically opposed. Depending on which aggression score was calculated (and reported) with the data from study 2, results could provide evidence that playing a violent digital game increases aggressive behavior, decreases it, or has no effect at all. The findings also suggest that volume and duration do not measure the same construct, although they clearly seem to be related. This does not necessarily constitute a problem with the CRTT. In fact, it could be considered a benefit if the CRTT was capable of capturing different (sub-) dimensions of aggressive behavior. However, no attempts to systematically disambiguate the latent variables supposedly measured by volume and duration have been made thus far.

These findings suggest that concerns about the CRTT'S standardization issues were justified. Of course, as the CRTT is the most common measure for aggressive behavior in the scholarly literature on violent game effects (C. Anderson et al., 2010), this has considerable implications. The results of studies that use the CRTT and meta-analyses that include these have to be interpreted with great caution. Moreover, given the questionable external validity of the test (Mitchell, 2012; Ritter & Eslea, 2005; Suris et al., 2004), researchers should be careful when they generalize results to situations outside the lab or make inferences about potential long-term effects to the point of public health issues. Of course, this issue is not limited to media effects research, as the CRTT is being used in a large variety of fields. This includes investigations of social and cerebral response in criminal psychopaths (Veit et al., 2010); effectiveness of prescription drugs in reducing hostility in panic disorders (Bond, Curran, Bruce, O'Sullivan, & Shine, 1995); and the facilitation of aggression through various substances, such as alcohol (Pihl et al., 1995). In some cases, practical recommendations for clinicians regarding the diagnosis (McCloskey, Berman, Noblett, & Coccaro, 2006) and treatment (Ben-Porath & Taylor, 2002) of patients are made based on results obtained with the CRTT. Given the impact of clinical research on the definition, assessment, diagnosis, and treatment of disorders in clinical practice, the importance of using objective, reliable, and valid measures cannot be overstated. The unstandardized use of the CRTT violates these requirements and, thus, poses a potential threat to the credibility of all laboratory research on aggressive behavior.

How to Advance a Field that is Loaded with Ideology

In view of the presented issues in theoretical conceptualizations, in manipulation and control of independent variables, and in the operationalization and measurement of dependent variables, it would not be sound to make any claims about conclusive evidence based on the available research. The conclusiveness of existing research on violent game effects is frequently overstated, and indulgence in ideological claims commonly go beyond what scientific evidence supports (Grimes et al., 2008). There appears to be a discrepancy between what media effects scholars find, and what some proclaim it means. Scholars have conjured violent games (and violent media in general) as a public health crisis, and claimed that it accounts for up to 30% of all violence in society (Strasburger, 2007), or that a strict ban of media violence would lead to an decrease of 10,000 homicides, 70,000 rapes, and 700,000 injurious assaults each year in the US alone (Centerwall, 1992). C. Anderson, Gentile, and Buckley (2007) consider violent video games as one of several risk factors that may cause aggressive, violent behavior and, in highly extreme and rare cases, even school shootings. Others draw rather curious comparisons, such as that the link between violent game use and aggression is as powerful as the link between condom use and prevention of HIV transmission, or as hazardous as smoking effects on lung cancer (C. Anderson et al., 2003). Not only does the alarmist manner in which a diffuse concept, such as aggression, is compared to a serious medical condition, such as cancer, unnecessarily heat the debate, it is also faulty on methodological grounds as the methodologies of media effects research and oncology are so drastically different that a comparison of the resulting effect sizes is invalid. If cancer studies would consist of participants smoking cigarettes for 5–10 min and then rating their cancer severity on a 5-point Likert scale or pushing a button when they recognize cancer-related words, then yes, such analogies would be eligible. But fortunately, cancer research does not have the methodology or validity issues that media effects studies do. Ironically, tests for cancer have everything that currently employed aggression tests do not. They are standardized, they are clinically validated (according to the results, one either has cancer or not), and they have a high reliability and external validity (someone who has cancer in a laboratory also has it outside the laboratory). Unfortunately, the same cannot be said about measurements of aggression.

There appears to be a discrepancy between what social scientists commonly measure in their laboratories and the behaviors that the public (or policy makers) are concerned about. Past research has usually not been conducted to inform public policy directly, but to advance academic knowledge of fundamental cognitive and behavioral processes in controlled laboratory environments. Consequently, when policy makers (e.g., Brown v. Entertainment Merchants Association, 2011) evaluated the empirical evidence, they did not find compelling proof of a link between media use and real-world violent behaviors – they could not, simply because the academic research, with few exceptions, has little bearing on societal violence. Unfortunately, scholars themselves are not always cautious, generalizing findings from weak laboratory studies to societal violence in ways that are inappropriate. The rhetoric to characterize these measures is exaggerated in the same way as the effects they ostensibly provide evidence for. Bushman and Gibson (2011), for example, describe the CRTT as "a weapon that could be used [by the participants] to blast their partner" (p. 30). Bushman and Huesmann (2014) compare the CRTT's noise blasts to the rock music played at excruciating volumes prisoners in Guantanamo Bay have been tortured with. To what extent this torture scenario, involving nonconsenting prisoners exposed to hours upon hours of sleep depriving noise, resembles the CRTT with its brief exposure and ostensibly consenting opponents in university laboratories remains unclear. Equating the CRTT to torture seems to be one more example of the irresponsible overreach to which this field has become accustomed.

The Disease of Moral Panic in Violent Games Research

But why is the public and scientific debate on violent games riddled with such a heated rhetoric? Offering one potential explanation, Gauntlett (2005) describes a phenomenon called moral panic. In a moral panic, a part of society considers certain behaviors or lifestyle choices of another part to be a significant threat to society as a whole, particularly when an older generation is not familiar with the behaviors of a younger generation (Kneer et al., 2012; Przybylski, 2014). In this environment, moral beliefs can substantially influence scientific research, and its results are readily used as confirmation for what has been suspected. Game researchers involved have a great interest in understanding the mechanisms of aggression to inform efforts at reduction of violent crime in society. Tackling an overt, proximal behavior, such as media use, has great merits: Attributing violence to manifest displays of media content that are considered immoral has convincing face validity. Moreover, media production and distribution could, in theory, be easily policed and regulated by state agencies. If media were causing harm in society, regulating them would be a fairly easy way of taking action against violent crime.

However, particularly when exaggerated, the danger of alarmist warnings about an overt, proximal behavior such as violent game use is a potential distraction from covert, distal issues rooted deep within society, such as poverty or inequality. Those problems are major sources of various societal issues, including violent crime, and are usually intangible, providing no ready 'bogeyman' in the parlance of moral panic theory – and are difficult issues to address. Just as testimony regarding the 'harmfulness' of comic books given to governments by mental health professionals in the 1950s now looks to be an example of a nannying excess on the part of the scientific community, so too, as Hall, Day, and Hall (2011) argue, will the extreme statements about effects of violent games do little other than to damage the credibility of the field. More than ten years ago, the journal *Nature* (2003) called on media violence researchers to "tone down the crusading rhetoric until we know more" (p. 355). Ten years later, we do know more, and what we know now does not suggest that it is time to return to crusading rhetoric. Far from it, it is increasingly time for the scientific community to employ a more cautious language and act as a voice of reason in the face of societal moral panics. It is imperative that the scientific community remains alert to these issues moving forward.

The Future of Game Violence Effects Research

Revisiting the four major issues this dissertation addresses, there is evidence for problems in (1) extant theories on the relationships between violent games and aggression, (2) the manipulation of independent variables and control of confounds as well as (3) the measurement of aggression in game violence experiments, and (4) ideological biases that shape both the research as well as the academic debate that surrounds it. Future research must tackle each of these problems to be able to determine whether a link between violent games and aggressive behaviors exist, and to inform the public about these results.

The future of violent game effects research needs testable theories predicting the role of game violence in aggressive behaviors. Biopsychosocial diathesis-stress approaches, such as the Catalyst Model (Ferguson, Rueda, et al., 2008), already account for how exposure to games might shape the individual ways violent crimes are ultimately committed. Given the relatively small role games play in the etiology of criminal behavior in this model, however, the psychological functions of game use are rather underdeveloped. Integrating these approaches to criminal behavior with motivational models of game use, such as mood management (Bowman & Tamborini, 2012; Zillmann, 1988) or uses-and-gratifications (Sherry, Lucas, Greenberg, & Lachlan, 2006), might be a viable solution. In contrast to socio-cognitive theories of aggression and media effects, these approaches are usually less concerned with passive learning through media contents, and more user-centric in explaining the functional link between psychological states and media exposure (Przybylski, Rigby, & Ryan, 2010). Such considerations are necessary when trying to explain why individuals may use the same media in very different ways, with very different outcomes (both intra- and interindividually). Thus, they could provide useful guidance when investigating whether specific use patterns of games, and not their contents, could be potentially detrimental to psychological well-being (including aggression).

On the empirical side, in light of the concerns about aggressive behavior or violent crimes precipitated by violent games, future studies should consider discontinuing investigations of game uses and effects in samples mostly consisting of college students. Studying game use patterns of offenders and those who have committed acts of violence against people or property instead could potentially yield highly interesting insights to our understanding of how and when violent media pose a risk. In addition, the identification of specific risk (and resilience) factors, such as an unfavorable family environment or mental health issues, preferably in prospective studies with actual control groups, might be an important future tasks for game violence researchers. Naturally, to be able to conduct these studies, the discussed problems in game effects research methodology have to be addressed first.

Methodological Rigor: A Potential Cure

While, as pointed out earlier, the propagation of extreme statements not supported by the available evidence is a problem of ideological convictions, the key condition enabling this current state of affairs are the insufficient or ambiguous methods employed to measure human aggression (Elson, Mohseni, et al., 2014; Ritter & Eslea, 2005), improper manipulations of conditions to test effects of game violence (Elson & Quandt, in press), or artificial situations under which games are studied (Elson & Breuer, 2014; D. Williams, 2005). With a corpus of precise and valid measurements for the different aspects of aggressiveness (thoughts, emotions, and behaviors), study results could no longer be subjected to interpretations from drastically different perspectives. In the case of research on the effects of violent digital games, the value of empirical evidence suffers greatly from the improper conclusions drawn based on results obtained through questionable methods. Accordingly, Elson and Ferguson (2014b) recommend scholars to adhere to two steps: First, not to generalize important findings further than the employed methods would allow (e.g., to consider aggression-related semantic activations simply as associations and not as ‘‘aggressive thoughts’’). Second, to overcome these limitations by developing standards to ensure objectivity and focus research on the proper validation of key measurements. There are current attempts to implement this, for example for the Hot Sauce Paradigm (Beier & Kutzner, 2012), and further investments in these directions should be encouraged.

In light of the recent replication crisis shaking up psychological science (Pashler & Harris, 2012), primarily research linking effects of cognitive priming with behavioral outcomes (Pashler, Coburn, & Harris, 2012), it became evident (more than ever) that ensuring the objectivity, reliability, and validity of research designs and key measures is paramount. This crisis has increased the awareness of scientific misconduct in relatively common questionable research practices (John, Loewenstein, & Prelec, 2012) and issues of undisclosed “methodological flexibility” (Simmons et al., 2011). It seems that “hot-button issues” in science are even more susceptible to these problems (Ioannidis, 2005), such as research on behaviors that large parts of the population engage in (e.g., playing digital games) or those that present a threat to societal values and norms (e.g., violence). Scientists are entrusted by the public to act as advocates when accumulated evidence is compelling. Yet, at the same time, they are obliged to be rather conservative and acknowledge the gaps and boundaries of scientific knowledge at any given point in time. The debate on whether or not playing violent games causes aggression or crime cannot be resolved simply because a large number of prominent scholars believe they do (as suggested by Bushman & Huesmann, 2014). One can clearly make an argument that games feature a large amount of morally objectionable content, and be offended by the excessive displays of violence in them. Others may object that whether or not violent content in games is repulsive, objectionable, or immoral might be relevant to policy makers, but not to researchers (Grimes et al., 2008).

But the scientific opinion of whether these contents lead to problematic behaviors in game players can and should only be formed through compelling methodologies that are able to produce a corpus of unambiguous findings. For the effects of violent games, however, this corpus can currently only be described as fragmentary, at best. Neither does the current state of research allow drawing the conclusion that violent games are *harmful*, nor does it allow inferring that they are completely *harmless* – simply because it is doubtful whether *harm* is actually being measured. From a scientific perspective, the development of improved methods and measures to close this gap is the key to overcome this problem. Yet what can be observed instead is that, as the results of certain studies reinforce the belief in harmful effects of violent games, some scholars have developed an ideological belief in the validity of the methods repeatedly employed in those studies as well. When it is argued that the empirical evidence on the link between game violence and aggression is not substantial enough to warrant definitive conclusions, responses usually point to the large number of experiments (in which the CRTT is very common) allegedly proving a causal relationship. The issue of lacking evidence for the external validity of those measures, which is necessary to make such a claim, is then refuted by claiming that the convergence of studies on media violence and aggression substantiates the validity of the measures commonly used (an example of this chain of arguments can be found in Bushman and Huesmann, 2014). This way, scholars develop a recursive argumentation in which a theoretical consideration demonstrates validity for methodological approach derived from it, and vice-versa.

This duality of ideologies, both on a theoretical and a methodological level, creates a vacuum in which science must necessarily stagnate. As Greenwald (2012) observed so keenly, “there is nothing so theoretical as a good method”, by which he argued that the multidecade durability of theory controversies in psychology can often be resolved through methodological advancements generating new data, which, in turn, can inspire novel theoretical considerations. The ideological rigidity in theory and methods that can be observed in violent game effects research, however, could stifle this synergy, as scholars try to find theoretical arguments why their methods are sufficient, and use the same methods to prove their theories were veritable in the first place. Whether or not the methodological insights presented in this dissertation (Elson, Mohseni, et al., 2014; Elson & Quandt, in press; Kneer et al., under review) and elsewhere (Adachi & Willoughby, 2011a; Ferguson & Savage, 2012; Järvelä, Ekman, Kivikangas, & Ravaja, 2014) will be able to ultimately overcome this impasse, however, remains speculative at this point. But as social-cognitive theories on violent games and aggression appear to be growing in their rigidity (Elson & Ferguson, 2014a, 2014b), especially in the face of an increasing number of failed replications, only methodological innovations can enable researchers to inform the public debate in a meaningful way, and enable the scientific field as a whole to advance as it should.

References

Adachi, P. J. C., & Willoughby, T. (2011a). The effect of violent video games on aggression: Is it more than just the violence? *Aggression and Violent Behavior*, *16*(1), 55–62. doi:10.1016/j.avb.2010.12.002

Adachi, P. J. C., & Willoughby, T. (2011b). The effect of video game competition and violence on aggressive behavior: Which characteristic has the greatest influence? *Psychology of Violence*, *1*(4), 259–274. doi:10.1037/a0024908

Anderson, C. A., Berkowitz, L., Donnerstein, E., Huesmann, L. R., Johnson, J. D., Linz, D., … Wartella, E. A. (2003). The influence of media violence on youth. *Psychological Science in the Public Interest*, *4*(3), 81–110. doi:10.1111/j.1529-1006.2003.pspi_1433.x

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*(1), 27–51. doi:10.1146/annurev.psych.53.100901.135231

Anderson, C. A., & Carnagey, N. L. (2009). Causal effects of violent sports video games on aggression: Is it competitiveness or violent content? *Journal of Experimental Social Psychology*, *45*(4), 731–739. doi:10.1016/j.jesp.2009.04.019

Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin, A. J., Eubanks, J., & Valentine, J. C. (2004). Violent video games: Specific effects of violent content on aggressive thoughts and behavior. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 36, pp. 199–249). New York, NY: Elsevier. doi:10.1016/S0065-2601(04)36004-1

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, *78*(4), 772–790. doi:10.1037/0022-3514.78.4.772

Anderson, C. A., Gentile, D. A., & Buckley, K. E. (2007). *Violent video game effects on children and adolescents: Theory, research, and public policy*. New York, NY: Oxford University Press.

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., … Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, *136*(2), 151–173. doi:10.1037/a0018251

Anderson, J. A. (2008). The production of media violence and aggression research: A cultural analysis. *American Behavioral Scientist*, *51*(8), 1260–1279. doi:10.1177/0002764207312019

Apperley, T. H. (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation & Gaming*, *37*(1), 6–23. doi:10.1177/1046878105282278

Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication*, *28*(3), 12–29. doi:10.1111/j.1460-2466.1978.tb01621.x

Barlett, C. P., Branch, O., Rodeheffer, C., & Harris, R. J. (2009). How long do the short-term violent video game effects last? *Aggressive Behavior*, *35*(3), 225–236. doi:10.1002/ab.20301

Baron, R. A., & Richardson, D. R. (1994). *Human aggression* (2nd ed.). New York, NY: Plenum Press.

Bartholow, B. D., & Anderson, C. A. (2002). Effects of violent video games on aggressive behavior: Potential sex differences. *Journal of Experimental Social Psychology*, *38*(3), 283–290. doi:10.1006/jesp.2001.1502

Bartholow, B. D., Bushman, B. J., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology*, *42*(4), 532–539. doi:10.1016/j.jesp.2005.08.006

Bartholow, B. D., Sestir, M. A., & Davis, E. B. (2005). Correlates and consequences of exposure to video game violence: Hostile personality, empathy, and aggressive behavior. *Personality and Social Psychology Bulletin*, *31*(11), 1573–1586. doi:10.1177/0146167205277205

Beier, R., & Kutzner, F. (2012). *Choose a juice! The effect of choice options and intention on aggression in a modified hot-sauce paradigm*. Paper presented at the 13th Annual Meeting of the Society for Personality and Social Psychology, San Diego, CA.

Ben-Porath, D. D., & Taylor, S. P. (2002). The effects of diazepam (valium) and aggressive disposition on human aggression. *Addictive Behaviors*, *27*(2), 167–177. doi:10.1016/S0306-4603(00)00175-1

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, *106*(1), 59–73. doi:10.1037/0033-2909.106.1.59

Bond, A., Curran, H. V., Bruce, M. S., O'Sullivan, G., & Shine, P. (1995). Behavioural aggression in panic disorder after 8 weeks' treatment with alprazolam. *Journal of Affective Disorders*, *35*(3), 117–123. doi:10.1016/0165-0327(95)00053-4

Bowman, N. D., & Tamborini, R. (2012). Task demand and mood repair: The intervention potential of computer games. *New Media & Society*, *14*(8), 1339–1357. doi:10.1177/1461444812450426

Boxer, P., Huesmann, L. R., Dubow, E. F., Landau, S. F., Gvirsman, S. D., Shikaki, K., & Ginges, J. (2013). Exposure to violence across the social ecosystem and the development of aggression: A test of ecological theory in the Israeli-Palestinian conflict. *Child Development*, *84*(1), 163–177. doi:10.1111/j.1467-8624.2012.01848.x

Breuer, J., Scharkow, M., & Quandt, T. (2013). Sore losers? A reexamination of the frustration-aggression hypothesis for colocated video game play. *Psychology of Popular Media Culture*. Advance online publication. doi:10.1037/ppm0000020

Brown v. Entertainment Merchants Association, 564 U.S. __ , 131 S.Ct. 2729 (2011).

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*(1), 219–229. doi:10.1037/0022-3514.75.1.219

Bushman, B. J., & Gibson, B. (2011). Violent video games cause an increase in aggression long after the game has been turned off. *Social Psychological and Personality Science*, *2*(1), 29–32. doi:10.1177/1948550610379506

Bushman, B. J., & Huesmann, L. R. (2014). Twenty-five years of research on violence in digital games and aggression: A reply to Elson & Ferguson (2013). *European Psychologist*, *19*(1), 47–55. doi:10.1027/1016-9040/a000164

Buss, A. H. (1961). *The psychology of aggression*. Hoboken, NJ: Wiley.

Carnagey, N. L., & Anderson, C. A. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition, and behavior. *Psychological Science*, *16*(11), 882–889. doi:10.1111/j.1467-9280.2005.01632.x

Centerwall, B. S. (1992). Television and violence. The scale of the problem and where to go from here. *Journal of the American Medical Association*, *267*(22), 3059–3063. doi:10.1001/jama.1992.03480220077031

DeWall, C. N., Finkel, E. J., Lambert, N. M., Slotter, E. B., Bodenhausen, G. V, Pond, R. S., … Fincham, F. D. (2013). The voodoo doll task: Introducing and validating a novel method for studying aggressive inclinations. *Aggressive Behavior*, *39*(6), 419–439. doi:10.1002/ab.21496

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. New Haven, CT: Yale University Press.

Elson, M., & Breuer, J. (2014). Isolated violence, isolated players, isolated aggression: The social realism of experimental research on digital games and aggression. In T. Quandt & S. Kröger (Eds.), *Multi.player. The Social Aspects of Digital Gaming* (pp. 226–233). London, UK: Routledge.

Elson, M., Breuer, J., Scharkow, M., & Quandt, T. (2014). *Digital games and frustration: Effects on aggression and cooperative behavior*. Paper presented at the 64th Annual Conference of the International Communicaton Association, Seattle, WA.

Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (2013). Comparing apples and oranges? Evidence for pace of action as a confound in research on digital games and aggression. *Psychology of Popular Media Culture*. Advance online publication. doi:10.1037/ppm0000010

Elson, M., & Ferguson, C. J. (2014a). Twenty-five years of research on violence in digital games and aggression: Empirical evidence, perspectives, and a debate gone astray. *European Psychologist*, *19*(1), 33–46. doi:10.1027/1016-9040/a000147

Elson, M., & Ferguson, C. J. (2014b). Does doing media violence research make one aggressive? The ideological rigidity of social cognitive theories of media violence and response to Bushman and Huesmann (2013), Krahé (2013), and Warburton (2013). *European Psychologist*, *19*(1), 68–75. doi:10.1027/1016-9040/a000185

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the Competitive Reaction

Time Task in aggression research. *Psychological Assessment*. Advance online publication. doi:10.1037/a0035569

Elson, M., & Quandt, T. (in press). Digital games in laboratory experiments: Controlling a complex stimulus through modding. *Psychology of Popular Media Culture*. doi:10.1037/ppm0000033

Ferguson, C. J. (2013a). *Adolescents, crime, and the media: A critical analysis*. New York, NY: Springer.

Ferguson, C. J. (2013b). Violent video games and the Supreme Court: Lessons for the scientific community in the wake of Brown v. Entertainment Merchants Association. *American Psychologist*, *68*(2), 57–74. doi:10.1037/a0030597

Ferguson, C. J. (2014). Violent video games don't exist. *Huffington Post*. Retrieved from http://www.huffingtonpost.com

Ferguson, C. J., & Dyck, D. (2012). Paradigm change in aggression research: The time has come to retire the General Aggression Model. *Aggression and Violent Behavior*, *17*(3), 220–228. doi:10.1016/j.avb.2012.02.007

Ferguson, C. J., Ivory, J. D., & Beaver, K. M. (2013). Genetic, maternal, school, intelligence, and media use predictors of adult criminality: A longitudinal test of the Catalyst Model in adolescence through early adulthood. *Journal of Aggression, Maltreatment & Trauma*, *22*(5), 447–460. doi:10.1080/10926771.2013.785457

Ferguson, C. J., & Rueda, S. M. (2009). Examining the validity of the modified Taylor competitive reaction time test of aggression. *Journal of Experimental Criminology*, *5*(2), 121–137. doi:10.1007/s11292-009-9069-5

Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S., & Smith, S. M. (2008). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation? *Criminal Justice and Behavior*, *35*(3), 311–332. doi:10.1177/0093854807311719

Ferguson, C. J., San Miguel, C., & Hartley, R. D. (2009). A multivariate analysis of youth violence and aggression: The influence of family, peers, depression, and media violence. *The Journal of Pediatrics*, *155*(6), 904–908. doi:10.1016/j.jpeds.2009.06.021

Ferguson, C. J., & Savage, J. (2012). Have recent studies addressed methodological issues raised by five decades of television violence research? A critical review. *Aggression and Violent Behavior*, *17*(2), 129–139. doi:10.1016/j.avb.2011.11.001

Ferguson, C. J., Smith, S. M., Miller-Stratton, H., Fritz, S., & Heinrich, E. (2008). Aggression in the laboratory: Problems with the validity of the modified Taylor competitive reaction time test as a measure of aggression in media violence studies. *Journal of Aggression, Maltreatment & Trauma*, *17*(1), 118–132. doi:10.1080/10926770802250678

Gauntlett, D. (2005). *Moving experiences: Media effects and beyond*. Eastleigh, UK: John Libbey.

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science*, *7*(2), 99–108. doi:10.1177/1745691611434210

Grimes, T., Anderson, J. A., & Bergen, L. A. (2008). *Media violence and aggression. Science and ideology*. Thousand Oaks, CA: Sage.

Hall, R. C. W., Day, T., & Hall, R. C. W. (2011). A plea for caution: Violent video games, the Supreme Court, and the role of science. *Mayo Clinic Proceedings*, *86*(4), 315–321. doi:10.4065/mcp.2010.0762

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, *13*(4), 339–363. doi:10.1080/15213269.2010.524912

Huesmann, L. R., & Skoric, M. M. (2003). Regulating media violence: Why, how and by whom? In E. L. Palmer & B. M. Young (Eds.), *The Faces of Televisual Media: Teaching, Violence, Selling to Children* (2nd ed., pp. 219–240). Mahwah, NJ: Lawrence Erlbaum Associates.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. doi:10.1371/journal.pmed.0020124

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*(3), 532–555. doi:10.1111/j.1460-2466.2007.00356.x

Järvelä, S., Ekman, I., Kivikangas, J. M., & Ravaja, N. (2014). A practical guide to using digital games as an experiment stimulus. *Transactions of the Digital Games Research Association*, *1*(2), 85–116.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. doi:10.1177/0956797611430953

Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological testing. Principles, applications, and issues* (7th ed.). Belmont, CA: Wadsworth.

Kneer, J., Elson, M., & Knapp, F. F. (under review). Fight fire with rainbows: The effects of displayed violence, difficulty, and performance in digital games on affect, aggression, and physiological arousal. *Journal of Media Psychology: Theories, Methods, and Applications*.

Kneer, J., Glock, S., Beskes, S., & Bente, G. (2012). Are digital games perceived as fun or danger? Supporting and suppressing different game-related concepts. *Cyberpsychology, Behavior, and Social Networking*, *15*(11), 604–609. doi:10.1089/cyber.2012.0171

Krahé, B. (2014). Restoring the spirit of fair play in the debate about violent video games: A comment on Elson and Ferguson (2013). *European Psychologist*, *19*(1), 56–59. doi:10.1027/1016-9040/a000165

Levendosky, A. A., Huth-Bocks, A. C., Semel, M. A., & Shapiro, D. L. (2002). Trauma symptoms in preschool-age children exposed to domestic violence. *Journal of Interpersonal Violence*, *17*(2), 150–164. doi:10.1177/0886260502017002003

Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, *25*(5), 331–348. doi:10.1002/(SICI)1098-2337(1999)25:5<331::AID-AB2>3.0.CO;2-1

Lindsay, J. J., & Anderson, C. A. (2000). From antecedent conditions to violent actions: A general affective aggression model. *Personality and Social Psychology Bulletin*, *26*(5), 533–547. doi:10.1177/0146167200267002

McCloskey, M. S., Berman, M. E., Noblett, K. L., & Coccaro, E. F. (2006). Intermittent explosive disorder-integrated research diagnostic criteria: Convergent and discriminant validity. *Journal of Psychiatric Research*, *40*(3), 231–242. doi:10.1016/j.jpsychires.2005.07.004

McMahan, R. P., Ragan, E. D., Leal, A., Beaton, R. J., & Bowman, D. A. (2011). Considerations for the use of commercial video games in controlled experiments. *Entertainment Computing*, *2*(1), 3–9. doi:10.1016/j.entcom.2011.03.002

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, *7*(2), 109–117. doi:10.1177/1745691611432343

Mussweiler, T., & Förster, J. (2000). The sex → aggression link: A perception-behavior dissociation. *Journal of Personality and Social Psychology*, *79*(4), 507–520. doi:10.1037/0022-3514.79.4.507

Nature. (2003). A calm view of video violence. *Nature*, *424*(6947), 355. doi:10.1038/424355a

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PloS ONE*, *7*(8), e42510. doi:10.1371/journal.pone.0042510

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531–536. doi:10.1177/1745691612463401

Pihl, R. O., Young, S. N., Harden, P., Plotnick, S., Chamberlain, B., & Ervin, F. R. (1995). Acute effect of altered tryptophan levels and alcohol on aggression in normal human males. *Psychopharmacology*, *119*(4), 353–360. doi:10.1007/BF02245849

Pinker, S. (2002). *The blank slate. The modern denial of human nature*. New York, NY: Penguin Books.

Przybylski, A. K. (2014). Who believes electronic games cause real world aggression? *Cyberpsychology, Behavior, and Social Networking*, *17*(4), 228–234. doi:10.1089/cyber.2013.0245

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*(2), 154–166. doi:10.1037/a0019440

Quandt, T., Breuer, J., Festl, R., & Scharkow, M. (2013). Digitale Spiele: Stabile Nutzung in einem dynamischen Markt. [Digital games: Stable patterns of use in a dynamic market.]. *Media Perspektiven*, *10*, 483–492.

Renaud, D., & Unz, D. (2006). Die M-DAS - eine modifizierte Version der Differentiellen Affekt Skala zur Erfassung von Emotionen bei der Mediennutzung [The M-DAS - A modified version of the differential affect scale to assess emotions through media use]. *Zeitschrift für Medienpsychologie*, *18*(2), 70–75. doi:10.1026/1617-6383.18.2.70

Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior*, *31*(5), 407–419. doi:10.1002/ab.20066

Scacchi, W. (2010). Mods, modders, modding, and the mod scene. *First Monday*, *15*(5). Retrieved from http://www.firstmonday.org

Sestir, M. A., & Bartholow, B. D. (2010). Violent and nonviolent video games produce opposing effects on aggressive and prosocial outcomes. *Journal of Experimental Social Psychology*, *46*(6), 934–942. doi:10.1016/j.jesp.2010.06.005

Sherry, J. L., Lucas, K., Greenberg, B. S., & Lachlan, K. A. (2006). Video game uses and gratifications as predictors of use and game preference. In *Playing Video Games. Motives, Responses, and Consequences* (pp. 213–224). Mahwah, NJ: Lawrence Erlbaum Associates.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. doi:10.1177/0956797611417632

Staude-Müller, 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*(1), 41–50. doi:10.1024/1421-0185.67.1.41

Strasburger, V. C. (2007). Go ahead punk, make my day: It's time for pediatricians to take action against media violence. *Pediatrics*, *119*(6), e1398–e1399. doi:10.1542/peds.2007-0083

Suris, A., Lind, L., Emmett, G., Borman, P. D., Kashner, M., & Barratt, E. S. (2004). Measures of aggressive behavior: Overview of clinical and research instruments. *Aggression and Violent Behavior*, *9*(2), 165–227. doi:10.1016/S1359-1789(03)00012-0

Tamborini, R., Weber, R., Bowman, N. D., Eden, A., & Skalski, P. (2013). "Violence is a many-splintered thing": The importance of realism, justification, and graphicness in understanding perceptions of and preferences for violent films and video games. *Projections*, *7*(1), 100–118. doi:10.3167/proj.2013.070108

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, *35*(2), 297–310. doi:10.1111/j.1467-6494.1967.tb01430.x

Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior*, *1*(2), 163–177. doi:10.1016/1359-1789(95)00014-3

Van der Dennen, J. M. G. (1980). *Doing violence to "aggression". Problems in the definitions and concepts of aggression, violence, and related terms*. Paper presented at the 4th Biennial Meeting of the International Society for Research on Aggression, Haren, The Netherlands.

Veit, R., Lotze, M., Sewing, S., Missenhardt, H., Gaber, T., & Birbaumer, N. (2010). Aberrant social and cerebral responding in a competitive reaction time paradigm in criminal psychopaths. *NeuroImage*, *49*(4), 3365–72. doi:10.1016/j.neuroimage.2009.11.040

Warburton, W. (2014). Apples, oranges and the burden of proof - Putting media violence findings in context. A comment on Elson and Ferguson (2013). *European Psychologist*, *19*(1), 60–67. doi:10.1027/1016-9040/a000166

Williams, D. (2005). Bridging the methodological divide in game research. *Simulation & Gaming*, *36*(4), 447–463. doi:10.1177/1046878105282275

Williams, K. D. (2009). The effects of frustration, violence, and trait hostility after playing a video game. *Mass Communication and Society*, *12*(3), 291–310. doi:10.1080/15205430802461087

Woodworth, R. S. (1938). *Experimental psychology*. New York, NY: Henry Holt.

Wright, J. C., Huston, A. C., Reitz, A. L., & Piemyat, S. (1994). Young children's perceptions of television reality: Determinants and developmental differences. *Developmental Psychology*, *30*(2), 229–239. doi:10.1037//0012-1649.30.2.229

Zillmann, D. (1988). Mood management: Using entertainment to full advantage. In L. Donohew, H. E. Sypher, & E. T. Higgins (Eds.), *Communication, Social Cognition, and Affect* (pp. 147–171). Hillsdale, NJ: Lawrence Erlbaum Associates.

Games and Software

Konami. (2003). Dance Dance Revolution Max 2. Tokyo, Japan: Konami.

Midway. (2004). Mortal Kombat: Deception. Chicago, IL: Midway.

Valve. (2007). Team Fortress 2. Bellevue, WA: Valve.

Appendix A

Publications within this cumulative dissertation and description of the doctoral candidate's contribution to each one.

1. Elson, M., & Ferguson, C. J. (2014). Twenty-five years of research on violence in digital games and aggression: Empirical evidence, perspectives, and a debate gone astray. *European Psychologist, 19*(1), 33–46. doi:10.1027/1016-9040/a000147

First author's contributions: Conceptualization of the article, composing theoretical considerations, review of the empirical literature, writing of the first draft, revising the manuscript after reviews.

Co-author's contributions: Assistance with the conceptualization, extension of theoretical review, review of meta-analytical work, writing of the first draft, assistance with revising the manuscript after reviews.

2. Elson, M., & Ferguson, C. J. (2014). Does doing media violence research make one aggressive? The ideological rigidity of social cognitive theories of media violence and response to Bushman and Huesmann (2013), Krahé (2013), and Warburton (2013). *European Psychologist, 19*(1), 68–75. doi:10.1027/1016-9040/a000185

First author's contributions: Conceptualization of the response, writing of the first draft, revising the manuscript after reviews.

Co-author's contributions: Assistance with the conceptualization, assistance with writing of the first draft, assistance with revising the manuscript after reviews.

3. Elson, M., & Quandt, T. (in press). Digital games in laboratory experiments: Controlling a complex stimulus through modding. *Psychology of Popular Media Culture*. doi: 10.1037/ppm0000033

First author's contributions: Conceptualization of the article, review of literature, data analysis, writing of the first draft, revising the manuscript after reviews.

Co-author's contributions: Proofreading and editing of the first draft and revised manuscript after reviews.

4. Kneer, J., Elson, M., & Knapp, F. F. (under review). Fight fire with rainbows: The effects of displayed violence, difficulty, and performance in digital games on affect, aggression, and physiological arousal. *Journal of Media Psychology: Theories, Methods, and Applications.*

Second author's contributions: Planning of the study, assistance in selection of methods and measures, assistance in data analysis, assistance in writing of the first draft.

Co-authors' contributions: Conceptualization of the article, data collection and analysis, writing of the first draft.

5. Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the Competitive Reaction Time Task in aggression research. *Psychological Assessment.* Advance online publication. doi:10.1037/a0035569

First author's contributions: Conceptualization of the article, literature review, data collection (study 2), standardization analyses for all three studies, writing of the first draft, revising the manuscript after reviews.

Co-authors' contributions: Assistance with the conceptualization, data collection (studies 1 and 3), reactivity analyses for studies 1 and 2, assistance with writing of the first draft, assistance with revising the manuscript after reviews.

Twenty-Five Years of Research on Violence in Digital Games and Aggression

Empirical Evidence, Perspectives, and a Debate Gone Astray

Malte Elson[1] and Christopher J. Ferguson[2]

[1]Department of Communication, University of Münster, Germany,
[2]Department of Psychology, Stetson University, DeLand, FL, USA

Abstract. Violence in digital games has been a source of controversy in the scientific community and general public. Over two decades of research have examined this issue. However, much of this research has been undercut by methodological limitations and ideological statements that go beyond what scientific evidence could support. We review 25 years of experimental, cross-sectional, longitudinal, and meta-analytical research in this field. Empirical evidence regarding the impact of violent digital games on player aggression is, at best, mixed and cannot support unambiguous claims that such games are harmful or represent a public health crisis. Rather, indulgence in such claims risked damage to the credibility of games effects research, credibility which can only be restored through better empirical research and more conservative and careful statements by scholars. We encourage the field to engage in a responsible dialog and constructive debate that could continue to be enriching and invigorating.

Keywords: digital games, mass media, aggression, violence

If expressed concerns about digital game violence as a cause of aggression and violent crimes were true, such as that they and other violent media are responsible for as much as 30% of societal violence (Strasburger, 2007), the implications would be extremely worrisome. Today, 1 in 4 Germans (Quandt, Scharkow, & Festl, 2010), 41% of Flemish residents (iLab.o, 2011), and more than half of the Finnish population (Karvinen & Mäyrä, 2011) consider themselves regular digital game players. As such, this is an important issue to consider, as much given the public perceptions of digital games and violent crimes as well as the practical implications of this argument. Similar to the public debate, a lively discussion has occurred within the scientific community about whether a causal or even correlational connection exists between digital game violence and real-life aggression and violence. This discussion has, at times, become polemic and, as Grimes, Anderson, and Bergen (2008) argue, the lines between objective science, politics, and advocacy have often become blurred. In this article we hope to examine this debate, particularly from a European perspective, and elucidate both the evidence for and against beliefs that digital games are involved in real-life violence, as well as the sociological processes that may have led to the scientific community speaking beyond the data available to support those causal beliefs. Instead of providing a definitive Yes or No answer regarding the impact of digital games on real-life aggression, we try to report findings from different perspectives on underlying explanatory mechanisms, consistencies and contradictions in empirical findings, views on the practical impact, and the role of digital game violence in society. Our goal is to present the state of the art of violent game effects research, but also what other variables might play a (more) important role on player behavior and need consideration in forthcoming studies. We also discuss "what went wrong" in the past scholarship, as we believe this is a necessary

DOI: 10.1027/1016-9040/a000147

precursor to improvement in the field and advance our understanding of the mechanisms behind violence in digital games.

The Debate on Digital Game Violence

At least historically, many researchers have been convinced of the detrimental effects of virtual violence (e.g., Anderson & Dill, 2000; Fischer, Aydin, Kastenmüller, Frey, & Fischer, 2012; Huesmann, 2010) particularly on player aggression. This was particularly true in the past decade when the field became dominated by advocates of the social-cognitive view of aggression, a theoretical model often closely tied to the "harm" position on digital games. In more recent years, however, there have been an increasing number of scholars (e.g., Ferguson & Kilburn, 2010; Sherry, 2007; Ward, 2011) who have expressed vocal skepticism of the "harm" view, or consider links between digital games and real-life aggression or violence to be weak or unimportant compared to other influences, especially in childhood and adolescence.

Historically, advocates of the "harm" view had taken to claiming that universal consensus existed to support their position. As early as 2003, some scholars declared that "the scientific debate over whether media violence increases aggression and violence is essentially over" (Anderson et al., 2003; p. 81). Despite this, debates within the scientific community have continued and only intensified in subsequent years. Even Huesmann's (2010) attempt of "(n)ailing the coffin shut on doubts that violent video games stimulate aggression" did not seem to have the desired outcome. How then is it possible that different researchers come to diametrically opposed conclusions about the state of the research when looking at the same published evidence?

Grimes et al. (2008) observe that the field of media violence is one example of wherein the politics, ideology, and personal beliefs on controversial societal issues fuel a heated scholarly debate. In this field, it had become common for scholars to assert rather extreme claims such as that the influence of digital games on aggression was similar in magnitude to the effect of smoking on lung cancer (Anderson et al., 2003), that a near universal consensus existed among scholars, or that the interactive nature of digital games made them more dangerous than other media. Grimes et al. note that in such an environment it is difficult to maintain scientific discourse on an objective examination of data and facts rather than a defense of rigid ideological positions. In the case of digital game violence, this has gone so far that instead of considering exclusively methodological rigor and validity of scientific research, some scholars supporting the "harm" view claim to have analyzed the expertise of scholars supporting and opposing California in Brown v. Entertainment Merchants Association (2011). Perhaps not surprisingly, they concluded that they and their colleagues must be considered "true aggression and violence experts" (Bushman & Anderson, 2011, p. 9), and those opposing California "are relatively unqualified to offer 'expert' opinions" (Sacks, Bushman, & Anderson, 2011, p. 12). However, independent scholars not involved in either amicus brief have already evaluated these claims and found them to be faulty on both methodological and theoretical grounds (Hall & Hall, 2011). To paraphrase Hall and Hall (2011), finding that one group of scholars compares themselves to their opponents and declares, by happy coincidence, that it is they and not their opponents who are the true experts is neither surprising nor illuminating. In fairness, once a debate becomes heated, both sides are likely to focus on refuting the other side rather than looking for ways to dialog and improve the science.

The risk, however, is the potential loss of credibility to the field (Hall, Day, & Hall, 2011). This potential ramification became particularly apparent when looking at how the scientific evidence had been perceived by courts and governments, such as the US Supreme Court in its Brown v. EMA (2011) decision. In this court case, the majority decision of the Supreme Court emphasized that the evidence presented by the state of California in its attempt to ban violent digital game sales to minors was not compelling. The court commented that the state had not presented studies showing a causal link between violent game playing and real-life acts of aggressiveness. Following the same rationale in an extensive literature review that expressed profound criticism of the existing evidence, the Australian attorney-general department (2010) decided to lift the ban on games exceeding the criteria for a 15+ rating. Similarly, a review of the evidence by the Swedish media council (Statens Medieråd, 2011) declared that the research evidence did not support links between digital games and real-world aggression and many of the existing studies, particularly those conducted in support of the "harm" position, were deeply flawed methodologically.

It is therefore commendable that scholars in the field have started to consider the complexity of digital games as stimuli (Ravaja & Kivikangas, 2009), to refine conceptual definitions of violence and/or aggression (Ferguson & Dyck, 2012; Grimes et al., 2008), and to promote more rigorous and effective research methods (Adachi & Willoughby, 2011a; Ferguson & Savage, 2012).

Explanatory Models and Theories

So far, two models explicitly trying to explain the role of violent games in real-life aggression have been published. Anderson and Bushman's (2002) General Aggression Model is based on several domain-specific social-cognitive theories, such as social learning, cognitive neoassociation, and excitation transfer. It has become the default model for many digital game researchers, particularly those who endorse the "harm" view when designing studies and interpreting results. Ferguson, Rueda, et al. (2008) took a different approach with their Catalyst Model, which is focused on biological determinants as well as the social context of family and peer groups. We discuss these models in some detail here.

General Aggression Model (GAM)

The GAM has its roots in social learning theory. The social learning theory of aggression (Bandura, 1978) explains that aggressive behavior is acquired by either direct experience or observation of attractive, rewarded models and subsequent imitation. Thus, new expectations about social mechanisms are developed, and old concepts are altered under frequent observation of certain behaviors. This approach explains how instrumental aggressive behaviors are understood and acquired, and beliefs about social behavior (e.g., hostility) are internalized. It is widely assumed that avatars in digital games can function as social models, and people can acquire knowledge structures and behaviors from them through in-game rewards (e.g., high scores) much as they learn from humans. Many scholars employing social learning theory suspect that games containing realistic violence that is not socially sanctioned within the game have a potentially strong detrimental effect on their users. Longer playing times also would facilitate this effect due to greater consolidation and reinforcement of the modeled behavior.

The basic assumption of the GAM (Anderson & Bushman, 2002) is that knowledge structures like perceptual and person schemata or behavioral scripts develop from experience and can influence (social) perception, behavior (conscious and automated), affect, and beliefs. The GAM focuses on episodes or "persons in situations." Situational features (e.g., aggressive cues, incentives) and/or personality variables (e.g., traits, beliefs, learned scripts) are considered *input variables*. Naturally, these features are highly interdependent. For example, increasingly violent persons might interpret ambiguous situations as more hostile than they actually are. The subsequent situational interpretation and behavioral intent is influenced by the current *internal state* that consists of cognition (e.g., hostile thoughts, scripts), affect (mood and emotion, expressive motor responses), and arousal. Resulting *outcome* behavior is dependent on either automatic or heavily controlled processes. Immediate appraisal and automatic (re-)action is relatively effortless and impulsive, occurring unconsciously and without requiring many cognitive resources. If a person has enough resources (mostly time and mental capacity) and the output is important, but the immediate appraisal is unsatisfying, the decision can be reappraised (numerous times, if necessary). In any case, the output determines a reaction, which becomes part of the input for the next episode. And in the long term, repeated episodes form more permanent perceptual, attitudinal, or behavioral patterns.

The Catalyst Model

The Catalyst Model of violent crime by Ferguson et al. (2008) focuses on innate motivations, biological dispositions, and other more fundamental environmental factors such as peer and family influences. The model states that an aggression-prone personality develops mostly through biological and genetic dispositions. However, these relatively invariant factors are moderated by environmental aspects (e.g., the family) in a positive or negative direction. Circumstantial short-term stressors or catalysts (e.g., financial difficulties, relationship problems) increase the likelihood for more aggressive behaviors in individuals with a relatively predisposed disposition toward aggression. Or put simply, biological factors combined with proximal social factors such as parental abuse or peer delinquency can make a person prone to aggressive behavior, but stress from the environment determines the motivation to do so. The likelihood to act aggressively or violently is increased in times where environmental stressors are plentiful or particularly prominent. Individuals with a high proneness to violence would naturally have a lower threshold to act aggressively, requiring fewer environmental stressors to motivate them, while others might have a relatively high tolerance for potentially stressing events.

The role of digital game violence in this model is not causal. Like other forms of media, digital games are considered potential stylistic catalysts, meaning that a person with a disposition for violence may act aggressively with similar "signature" elements to actions seen in a digital game. The way in which violent behaviors are expressed specifically may be influenced by violent media (e.g., wearing the same clothes like a violent media character), but not be the reason or motivation to act violently in the first place. The acts of violence would still occur in another form, even without previous exposure to violent games. An individual with a disposition for violence would be susceptible to violence even when presented with contrasting modeling opportunities. However, individuals with an aggressive personality would be more attracted to violent media.

Strengths and Weaknesses

The distinct strength of the GAM lies within its unification of several social-cognitive learning theories, cognitive association processes, and moderators like physiological arousal. Theoretical foundations are combined into a simple model that aims to predict antecedents and consequences of human aggression. It includes many social-cognitive factors and thus provides a comprehensive social-cognitive research framework, making it the "default model" in media effects research (at least historically). However, the simplicity is a double-edged sword, and comes at a price. Concerns are raised such as that it overfocuses on cognitive scripts and does little to elucidate affective or personological variables that may influence aggression, thereby rendering the GAM a "tabula rasa" theory (Pinker, 2002) in function if not form. Interestingly, although social-cognitive learning processes are the hub of the GAM, it only marginally accounts for competing cognitive schemata. Even when considering that aggressive schemata and scripts are learned by playing violent games, people tend to do other things as well, and subsequently acquire different or contradicting schemata. It is also questionable how media are understood within the GAM, since they are considered to be equally capable of modeling aggressive behavior as actual incidents of aggression (e.g., within a family). One further criticism of the GAM is that the personological and biological inputs (and immunizing factors) are so underdeveloped in the model that they function as "fig leaves" to mask what is, in effect, little more than a basic script theory of aggression (Ferguson & Dyck, 2012). Also, it is not actually used by clinicians or other professionals in the field of pathological aggression (Ferguson & Dyck, 2012). We understand that the GAM is commonly used when researching social-cognitive processes of aggression, but are skeptical about its use in predicting media effects.

Contrary to the GAM, the Catalyst model considers individuals "active" modelers of their own behavior, so they seek out modeling opportunities according to the innate motivational system. Individuals with a predisposition for violence would try to seek models from which violent behaviors could be learned, and they would still be prone to act aggressively when presented with contrasting modeling opportunities. Similarly, less susceptible individuals would try to find nonviolent behavior modeling opportunities and be resistant to adverse models. However, the Catalyst model is a relatively new theoretical model of violent behavior and has received only little attention compared to the GAM.

Empirical Evidence

Empirical research on adversarial effects of digital games can be divided into three categories: experimental and causal studies, cross-sectional correlation and longitudinal studies, and meta-analyses. In the following section, we will present different empirical findings regarding the effects of digital violence on aggressive cognitions, emotions, and behaviors, and the methods used to assess them. Our aim is to give an exhaustive review of results obtained in laboratories and the field, to integrate different perspectives and interpretations, and to explain their relevance to the understanding of media effects.

Experimental and Causal Studies

The main body of psychological research on the effects of digital games consists of laboratory experiments. Many of these studies share a certain design: Study participants (mostly college students, often psychology or communication majors) either play a violent (mostly a first-person shooter) or a nonviolent game. Physiological arousal (heart rate, skin conductance level) is sometimes measured simultaneously, or before and after play. Afterwards, participants perform a test or fill out a questionnaire to assess aggressive cognitions, emotions, or behaviors, which are then compared for the two groups.

Aggressive Cognitions

There are a number of studies that investigate the facilitation of aggressive cognitions (e.g., thoughts) through violent digital game playing. While cognitions themselves are difficult, if not impossible to assess, there are superficial features of actual cognitions like semantic activation or accessibility of aggression-related concepts that are relatively easy to measure. However, the mere accessibility is not problematic, as it does not consequently result in any form of intent, let alone behavior. We would not reasonably conclude that having such associations leads one to intend to commit aggression or violent crimes or go to war any more than being primed with an image of whiskers would lead one to intend to be a cat. In fact, it would probably be evidence of neuropsychological impairment if a stimulus did not cause any associations with related constructs in a person. Still, measuring aggressive cognitions can help us in understanding how players experience (violent) games.

One popular way to measure accessibility of aggressive thoughts is the word completion task, which involves filling in one or more missing letters in a list of ambiguous items that can make more than one word (e.g., "explo_e" having the two possible completions "explore" or "explode"). An "aggressive cognition" score is then calculated for each participant by dividing the number of aggressive word completions by the total number of completions. This measurement has been used by several authors with significant results, indicating that playing violent digital games facilitates the accessibility of aggressive thoughts (Anderson et al., 2004; Barlett, Branch, Rodeheffer, & Harris, 2009; Barlett & Rodeheffer, 2009; Carnagey & Anderson, 2005; Sestir & Bartholow, 2010), although Cicchirillo and Chory-Assad (2005) did not find any such differences between experimental groups.

There are other methods to measure the accessibility of aggressive thoughts: Anderson and Carnagey (2009) found that playing a violent game led to shorter reaction times between on-screen presentation and verbal identification of aggressive words (e.g., assault, choke) compared to playing a nonviolent one. This effect was largely moderated by high trait aggression, however. Similarly, Giumetti and Markey (2007) found that only participants with a high dispositional anger in a violent-game condition gave more aggressive responses when they were asked to write down 20 unique things that protagonists of short stories with negative outcomes might do, feel, or think. Using the same method, Hasan, Bègue, and Bushman (2012) replicated the game violence effect, although unfortunately they did not measure their participants' trait aggressiveness. Another method was used by Ivory and Kalyanaraman (2007), who let their participants rate the similarity of aggressive (e.g., choke, wound) and ambiguous (e.g., animal, drugs) word pairs. A higher accessibility of aggressive thoughts should have led to relatively more aggressive interpretation of ambiguous words, resulting in higher similarity ratings. However, the test did not yield any significant results between the experimental groups.

As a measure for hostile perception, Brady and Matthews (2006) showed their participants a video in which a teacher asks a student to speak with him at the end of class, and rated the likelihood that the teacher would accuse the student of cheating. However, the authors did not find a difference between a high violent and a low violent game group. Focusing on implicit associations of aggressive cognitions with the self, Uhlmann and Swanson (2004) measured the effects of violent games on implicit self-concept with the implicit association task using the focal categories "aggressive" and "peaceful" on the target categories "self" and "other." Playing a violent game leads to shorter reaction latencies on the "self = aggressive" than the "self = peaceful" tasks (this finding was replicated by Bluemke, Friedrich, and Zumbach, 2010). Thus, although there are some inconsistencies in the research, many studies suggest that people who have just played a violent video game subsequently have more aggression-related associations than people who played another, nonviolent game. This sort of finding was described as "common sense" by the US Supreme Court, noting (correctly) that there was no evidence such cognitions led to intent, let alone behavior. Indeed the very use of the term "aggressive thoughts or cognitions" may be disingenuous to the degree they conflate intents or cognitive hostility with priming of cognitive associations.

There are, however, other factors besides just displayed violence to be considered when measuring accessibility of aggressive thoughts. Schmierbach's (2010) study on the mode of play indicates that playing cooperatively leads to a significantly lower accessibility of aggressive thoughts compared to playing competitively or solo. Moreover, the gender of the opponent has an influence on aggressive thoughts as well (Eastin, 2006). The importance of considering motivational aspects in digital game effects was underlined by Denzler, Häfner, and Förster (2011) who found that when playing a violent game with the goal to vent anger, accessibility of aggressive thoughts (measured with lexical decision tasks) was actually *inhibited*. Kneer, Munko, Glock, and Bente (2012) showed that young adults suppressed aggressive concepts when primed with violent game content as an implicit defense mechanism for their own gaming habits or even those of the generation they belong to (see also Kneer, Glock, Beskes, & Bente, 2012). Thus, many investigations found that violent content in games increases the accessibility of aggressive thoughts. However, this effect appears to be far more context specific than had previously been indicated and is mitigated or even inverted by many internal and external variables.

Aggressive Emotions

A large body of research on effects of violence in games deals with aggressive affect like anger or hostility, usually by means of participants' self-reports or rather distal physiological indices. Arriaga, Esteves, Carneiro, and Monteiro (2006) used the State Hostility Scale (SHS; Anderson, Deuser, & DeNeve, 1995) to describe the participants' current aggressive feelings, using ratings of 35 items, yielding significantly higher hostile feelings for participants who played a violent game compared to a nonviolent one. This finding is consistent with some other research reports (e.g., Barlett et al., 2009; Carnagey & Anderson, 2005; Saleem, Anderson, & Gentile, 2012; Sestir & Bartholow, 2010), but there are quite a few who found mixed evidence (Anderson & Carnagey, 2009), or no effects (Ballard, Hamby, Panee, & Nivens, 2006; Ferguson & Rueda, 2010; Ivory & Kalyanaraman, 2007; Valadez & Ferguson, 2012).

There has been some more content-specific research, investigating particular properties of displayed violence. Barlett, Harris, and Bruey (2008) found a significant increase in hostility when moderate and high amounts of blood were visible, but not with low or no blood. The results of Jeong, Biocca, and Bohil (2012), however, suggest that this effect might be fully mediated by spatial presence. In another study, presence of blood did not have any effect on hostility (Farrar, Krcmar, & Nowak, 2006). Barlett and Rodeheffer (2009) investigated the effects of realism in digital games, and found that participants who played a realistic violent game had a higher SHS score than those who played an unrealistic violent or nonviolent game. Eastin (2007) also showed that group size and game mode (cooperative vs. competitive) might be confounding factors to consider when measuring effects of displayed violence. Further research has been conducted on other negative emotions (sometimes linked to aggression, and sometimes not). Brady and Matthews (2006) found that playing a highly violent game (compared to a less violent one) increased negative emotions in general, while Baldaro et al. (2004) found no significant effect on physical aggressiveness, indirect hostility, irritability, negativism, resentment, suspiciousness, verbal hostility, or feelings of guilt. In another study, participants reported more positive attitudes toward traffic delinquency using a delinquency-reinforcing game, while there was no effect on aggressive emotions (Fischer et al., 2012). In-game justification of the digital violence also seems to matter, as participants in the study of Hartmann, Toz, and Brandon (2010) felt guiltier when their violent actions were presented as unjustified (see also Hartmann & Vorderer, 2010).

While there are already considerable limitations of using self-report data to assert temporary changes in aggression affect, even more methodological issues are introduced by employing measurements specifically designed to measure trait aggression. Uhlmann and Swanson (2004) measured postgame trait aggression without any significant findings regarding violent content. By contrast (and quite puzzlingly given the presumed consistency of trait aggression), Frindte and Obwexer (2003) observed changes in trait aggression after violent game play, but not in state anger.

Unsworth, Devilly, and Ward (2007) were concerned with the overall generalizability of violent game effects on aggressive feelings in studies like the ones cited above. They found that the significant effect on state anger in their sample was actually caused by a small subsample of 1.87% that had a clinically relevant aggression score, while the main body of participants remained unaffected (or in some cases even experienced a decrease in anger). This suggests that only a very small part of the population could be prone to possibly detrimental effects of game violence. As such the body of work on aggressive emotions presents a complex array of significant and null studies. Many of the studies find inconsistent and often opposing results. Overall, results linking violent digital games with aggressive affect were less consistent and yielded smaller effects than for aggressive cognitions. Such studies were also often impaired by high potential for demand characteristics achieved through presenting independent and dependent variables very close temporally and using highly obvious measures of aggressive affect (i.e., having participants play a violent game, then asking them if they feelings of aggression). Given the issues in measuring a complex variable like aggressive emotions, and the many context variables that appear to be important but are often not considered, the results in this area are overall fairly inconclusive.

Aggressive Behavior

Even if violent digital games consistently caused an increase in aggressive semantic activations and affect across studies, most of the discussion of potential "harm" within the scientific community, news media, and the general public focused on the issue of whether violent digital game exposure results in aggressive or violent actions. However, this has been a difficult question to answer. Legal and ethical restrictions make measuring aggressive behavior in a laboratory a difficult enterprise. As can be imagined, it is generally not possible to create a scenario in which individuals will attack each other in the laboratory environment. Unfortunately, this means that most experiments must rely on instruments do *not* measure aggression, but vaguely approximate it in some way. An instrument used in many experimental studies is the Competitive Reaction Time Task (CRTT, originally by Taylor, 1967), in which the participants play a number of trials of a reaction time game against a (fictional) opponent and the loser of each trial gets punished by the winner. In digital games studies, the electric shocks that Taylor used as punishment have been replaced with noise blasts, whose intensity and/or duration (the measure for aggressiveness) can be varied by the participant. While this test has received a lot of criticism for its lack of standardization and validity (Ferguson, Smith, Miller-Stratton, Fritz, & Heinrich, 2008; Savage, 2004; Tedeschi & Quigley, 1996), it is still widely being used. Using at least 13 different modifications to the CRTT's procedure or raw score analysis, several authors have found that playing a violent game compared to a nonviolent resulted in higher CRTT scores (Anderson & Carnagey, 2009; Anderson et al., 2004; Bartholow & Anderson, 2002; Bartholow, Bushman, & Sestir, 2006; Bartholow, Sestir, & Davis, 2005; Carnagey & Anderson, 2005; Konijn, Nije Bijvank, & Bushman, 2007; Sestir & Bartholow, 2010), while others found mixed evidence (Anderson & Dill, 2000; Arriaga, Monteiro, & Esteves, 2011), or no effects at all (Ferguson & Rueda, 2010; Ferguson, Rueda, et al., 2008; Elson, Breuer, Van Looy, & Kneer, 2012).

However, the implications of the results gathered with the CRTT are diminished by its methodological flexibility, as the lack of standardization in test procedure and data analysis breeds problems for the test's objectivity (Breuer, Elson, Mohseni, & Scharkow, 2012). Unstandardized testing and processing of raw data yield unstandardized test scores, thus constraining the test's approximation to the true value (aggressive behavior), and making it difficult to compare studies that used the test differently (e.g., in meta-analyses). It remains puzzling to us why so many different versions exist and the field has resisted agreement on a standardized measurement technique for this measure. Furthermore, the CRTT does not appear to predict real-world aggression (Ferguson & Rueda, 2009) nor is it influenced by actual habitual media violence use in real life as would be expected by the "harm" view (Krahé et al., 2011). Nor, despite being quick, easy to use, and freely available, is the CRTT used to predict aggression in clinical settings. Again, we strongly believe that researching causes and antecedents of aggression is a highly relevant undertaking. However, the problems associated with the CRTT, at least as it is currently being used, are constraining the credibility and significance of laboratory research on human aggression.

Another laboratory measure used for aggressive behavior is the Hot Sauce Paradigm (Lieberman, Solomon, Greenberg, & McGregor, 1999), in which aggressiveness is measured by the amount of hot sauce that participants use to prepare a cup of chili sauce for another (fictional) participant. Some studies found that playing violent games leads participants to use more hot sauce to spice the chili (Barlett et al., 2009; Fischer, Kastenmüller, & Greitemeyer, 2010), while Adachi and Willoughby (2011b) demonstrated that this is likely caused by a game's competitiveness, not its violent content. Like the CRTT, however, the validity of the Hot Sauce paradigm has been questioned (Ritter & Eslea, 2005). A main issue with measures such as the CRTT and Hot Sauce Paradigm is not only their unstandardized use but their generalizability to real-world aggression. Naturally, children (and adults) wishing to be aggressive do not chase after their targets with jars of hot sauce or headphones with which to administer bursts of white noise.

Other researchers were interested in hostile or mildly delinquent behaviors rather than aggression. Participants in the study of Fischer et al. (2012) were likelier to steal pens or candy bars from the laboratory after playing a delinquency-reinforcing game compared to a delinquency-neutral game. Using a similar procedure, this finding was replicated by Happ, Melzer, and Steffgen (2011), who also observed that playing a violent game leads to less prosocial behavior, which was assessed by the willingness to fill out an optional questionnaire. Greitemeyer and McLatchie (2011) showed that another participant's (a confederate) job-relevant qualifications were evaluated less positively after playing a violent game. Similar to the measures discussed above, however, there is a lack of evidence to which situations or behaviors these results might be generalized.

A large body of research does not focus on aggressive behavior, but instead measures related constructs like cooperativeness (usually with "mixed-motive" games).

For example, Brady and Matthews (2006) showed that displayed violence led to more uncooperative behavior in a game with another (fictional) participant. The results of Rothmund, Gollwitzer, and Klimmt (2011) suggest that cooperativeness is diminished in particular when the in-game violence is perceived from a victim's perspective. However, using a similar decision dilemma, Greitemeyer, Traut-Mattausch, and Osswald (2012) show that playing a violent game cooperatively leads to more postgame cooperative behavior compared to playing alone, or playing a neutral game. This result is corroborated by two other recent studies that found, compared to competitive play, playing violent games cooperatively increases helping behavior (Ewoldsen et al., 2012; Velez, Mahood, Ewoldsen, & Moyer-Gusé, 2012). Recent work by Jerabeck and Ferguson (2012) found that playing violent games had no influence on either aggressive behavior or prosocial behavior. However playing video games cooperatively, whether violent or nonviolent, increased cooperative behavior. A recent Swedish study (Bennerstedt, Ivarsson, & Linderoth, 2012) found that players actually *increased* their cooperative behaviors while playing violent video games. The authors further concluded that many past studies had made serious errors in setting up artificial scenarios rather than examining more closely the experience of gamers.

As such, the body of research on the link of violent games and aggressive behavior is inconsistent. Many studies pointing to such an effect suffer from weak methodologies and an artificial setup of both the measures and the playing situation itself, while more carefully designed experiments show there are many variables to be considered that are more important than violent content. This regards characteristic features in game design besides violence that need to be considered (e.g., competitiveness), as well as playing modes (competitively vs. cooperatively), and contextual variables (e.g., playing against a friend vs. the computer). Without proper stimulus selection and experimental control of other variables, the current evidence does not provide the consistent results necessary to resolve this controversial debate. Further, experiments employing standardized outcome measures are less likely to demonstrate negative effects than those employing unstandardized measures (Ferguson & Kilburn, 2009) giving credence to the methodological flexibility issue. Media effects research requires standardized and validated instruments in order to come to consistent and convincing conclusions.

Cross-Sectional Correlation and Longitudinal Studies

With the growing body of evidence from experimental studies, which had remained inconsistent in outcome, there has been a stronger demand for longitudinal work to determine whether exposure to violent digital games could lead to long-term negative outcomes. Many researchers have noted the lack of an observable impact of violent digital games on actual crime rates, as there seems to be a negative relation between the spread of digital games and violent delinquency over the last decades (Ferguson, 2010; Sherry, 2007; Ward, 2011). Although the considerable declines in youth and adult violence cannot be attributed to the proliferation of violent games (such would be an ecological fallacy), it is nonetheless a compelling piece of evidence demonstrating extreme claims (e.g., stating the risks of violent games are greater than parental abuse) are simply nonsensical.

Unfortunately, there is yet no standardized instrument to assess violent game exposure. Most researchers tend to use some variant of the Violent Video Game Exposure (VVGE) questionnaire first introduced by Anderson and Dill (2000), in which participants' playing frequency and violent content ratings of their five favorite games are multiplied and averaged to form a composite score. Recent work has suggested such approaches to media violence exposure may not be accurate in representing adolescent's actual exposure to violent content and may spuriously inflate effect size estimates (Fikkers, Valkenburg, & Vossen, 2012). Several studies have linked VVGE to self-reported aggression-related variables, such as delinquency (Anderson & Dill, 2000), physical and verbal aggression (Anderson et al., 2004; Bartholow et al., 2005), or anger (Koglin, Witthöft, & Petermann, 2009), while other studies have not found a correlation between VVGE and trait aggression (Ferguson & Rueda, 2009; Ferguson, Rueda, et al., 2008; Ferguson, San Miguel, & Hartley, 2009; Puri & Pugliese, 2012), youth violence (Ferguson, 2011; Gunter & Daly, 2012; von Salisch, Vogelgesang, Kristen, & Oppl, 2011), or attitudes toward violence (Brady & Matthews, 2006). Ferguson and Rueda (2010) even found that participants with a high VVGE had a significantly reduced state hostility after a stressful task. This is corroborated by the results of Puri and Pugliese (2012) who found that use of digital role-playing games (that include violence) was negatively related to aggression. However, there is evidence that any link between VVGE and aggressive behavior is largely mediated by other variables, such as hostile expectations, beliefs about aggression, or arousal (Barlett et al., 2009; Zhen, Xie, Zhang, Wang, & Li, 2011). Then again, Gunter and Daly (2012) show that any correlation between VVGE and self-reported delinquency in an unmatched sample was turned to nonsignificance when the sample was matched using propensity scores. In a survey of correctional inmates, Surette (2012) found weak evidence for violent game effects, but stronger evidence for their function as stylistic catalysts. Ferguson and Garza's (2011) results even show that exposure to action games interacted with parental involvement to increase the likelihood of volunteering for civic engagement.

In the 2-year prospective study of German adolescents by Hopf, Huber, and Weiß (2008), VVGE at time 1 was a significant (yet small) predictor of aggressive behavior and delinquency at time 2. Möller and Krahé (2009) found similar effects of violent game playing on physical aggression in a sample of German students 30 months later; with mediators and moderators like hostile attribution and normative beliefs taken into account, however, this effect was reduced to nonsignificance. In an annual survey of US adolescents over 3 years, Willoughby, Adachi, and

Good (2011) showed that sustained violent game play was significantly related to steeper increases in US adolescents' trajectory of aggressive behavior, yielding a small effect even when taking numerous covariates (such as demographic variables, academic performance, peer deviance, or parental relationship) into account. However, they concluded this may be due to competitive gaming rather than violent content. Using three samples from Japan and the US, Anderson et al. (2008) found a weak link between VVGE and physical aggression (assessed with a different measure in each sample, ranging from a 1-item self-report scale to teacher and peer reports) 4 to 6 months later. In most of these prospective analyses (with the exception of Willoughby et al., 2011) little effort was made to control for other important risk factors for youth aggression such as family, peer, and personality factors. Furthermore, in all of these studies, the outcome measures used were not well-validated clinical measures of aggression.

There are other prospective and longitudinal studies which do not support a direct link between violent game exposure and aggressiveness over time. In a Finnish adolescent sample, Wallenius and Punamäki (2008) did not find violent game playing to be a significant predictor of direct aggression 2 years later when controlling for the potential confounding variables such as sex, age, and parent-child communication. In a Hispanic US sample, using ESRB ratings for violent content instead of self-report, Ferguson (2011) did not find a relationship between game playing at time 1 and aggression or delinquency at time 2 one year later. A 3-year longitudinal study with a sample from the same population yielded no effects of violent games on delinquency, aggressiveness, or dating violence (Ferguson, San Miguel, Garza, & Jerabeck, 2012). A recent publication from a second sample of Hispanic youth likewise found no evidence for a link between violent digital game exposure and youth violence, bullying, or a reduction in civic or prosocial behaviors 1 year later (Ferguson, Garza, Jerabeck, Ramos, & Galindo, 2013). Instead of the standard self-report measures, von Salisch et al. (2011) used expert ratings of digital game violence, as well as peer and teacher nominations for aggressive behavior. Taking into account several important third variables, they did not find game violence exposure to increase aggressiveness in a 1-year cross-lagged panel study. However, the authors found a considerable preference in participants with a high aggressiveness at time 1 to play violent games at time 2, a selection effect likely to skew results in correlational and longitudinal studies when not controlled for.

These longitudinal studies have generally been more effective in controlling for other important risk factors and using well-validated clinical measures of aggression, bullying, and violence. Although the overall evidence is, again, mixed, we conclude that studies which use more careful methodologies are least likely to find negative effects. Longitudinal work has been useful in identifying the mechanisms behind the link of aggressive personalities and violent media use often obtained in correlational studies, as there seems to be strong evidence for a selection effect. Due to the lack of proper variable control, there is little empirical evidence for whether specific situational or personological variables might exist that would foster the selection of violent games as a risk factor for aggressiveness or other detrimental behaviors.

One issue people are particularly worried about is the impact of digital games on children and adolescents. Therefore, longitudinal studies are particularly helpful when considering developmental effects of media use. Von Salisch et al. (2011), for example, suspect that the strong selection they found could be the beginning of a downward spiral (see also Slater, Henry, Swaim, & Anderson, 2003), in which problematic behaviors would manifest only in later developmental stages, particularly adolescence. As this is, however, not supported by their own data on 9- to 13-year-old children, further research would need to identify the stage in which selection effects would turn into a reciprocal behavior effect. There are other longitudinal studies (e.g., Ferguson, 2011; Wallenius & Punamäki, 2008) that do not provide evidence for this hypothesis. Adachi and Willoughby (2012) lament that research on positive outcomes of game playing is relatively neglected compared to the vast amount of deficit-oriented studies. They conclude that digital games may facilitate positive youth development, and call for further research to determine when games might be a serious risk factor, and when they might benefit children and adolescents.

Meta-Analyses

In spite of the debate about conceptual and methodological issues in game violence research, several authors tried to summarize the primary experimental and correlational data into meta-data, and to determine the overall effects on all aspects of aggression. Anderson et al. (2010) found a total of 136 published studies and found overall small effects (ranging from $r = .07$ to .21) for all of GAM's aggression components (cognitions, affect, behavior, and arousal). Effects for longitudinal studies, in particular, were negligible with $r = .075$ when controlling for time 1 aggression (but no other of the many relevant variables). They also categorized studies according to their methodological rigor with a criteria catalog by the authors, and found that research with "best practice" finds stronger results compared to "not best practice." Unfortunately, this coding guide is described only rather vaguely, and – for the lack of a clear definition – some points have been left entirely to their subjectivity (e.g., "the violent game contained little or no violence"). We would also suggest extending the catalog by some specific critical coding issues, namely the misuse of unstandardized aggression measures (such as the CRTT). Further, the authors included many of their own unpublished studies and those of close colleagues, but they did not solicit unpublished studies from authors whose work differed in results from their own, thus setting up selection bias problems (Ferguson & Kilburn, 2010). Thus, even though reporting mostly weak links, the Anderson et al. (2010) meta-analysis has to be considered with reasonable caution.

The meta-analytical work of Sherry (2001, 2007) yielded somewhat weaker effects for the overall link between violent game playing and aggression (r = .15), and also showed that survey studies and paper-pencil measures tend to produce larger effects than experimental studies and behavioral measures. Sherry specifically questions the practical implications of these results and dismisses the alleged observable impact of digital violence in society. The meta-analysis of Ferguson and Kilburn (2009) strengthens this observation, yielding effect sizes of a similar magnitude. However, they also accounted for the presence of a publication bias in the literature, resulting in a marginal effect size of r = .08. Both Ferguson and Kilburn (2009) and Sherry (2001, 2007) have rejected the view that the data supports a link between violent digital games and aggression. They consider the possibility of finding effects in the controlled environment of a laboratory, but express doubts about a notable impact of those small effects in real life. Ferguson and Kilburn (2009) also reject the idea that these small effects could be additive over time, as longitudinal studies usually find the weakest evidence that violent games increase aggressive behavior. Both Sherry (2001) and Ferguson and Kilburn (2009) also find evidence that mean effect sizes in meta-analyses are likely inflated due to weak methodology, the use of unstandardized outcome measures (i.e., methodological flexibility), and publication bias. As such, the ''true'' effect of video games is probably smaller than the mean effect found in any of the meta-analyses. Thus the conclusions of Anderson et al. (2010) are not replicated by other meta-analyses.

Moving the Debate Forward

Violence in digital games has been the center of several decades of research, as well as considerable controversy regarding the meaningfulness of that research. In the current review we agree with the recent assessments of the US Supreme Court (Brown v. Entertainment Merchants Assn., 2011), Australian Government (2010), and Swedish Government (2011) that the research has been inconsistent, and often besotted with serious methodological limitations. Furthermore, we agree with Hall et al. (2011) that peer review in this field has been insufficient, allowing for the proliferation of extreme statements that went beyond the available data, ultimately damaging the field's credibility. We lament this state of affairs, although we acknowledge that considerable strife between diametrically opposed positions is natural during a period of paradigm change. And although we cannot agree with the statements made by some scholars on this issue, we recognize that these statements were made in good faith. Further, we do also acknowledge that some scholars, who have advocated for the ''harm'' position (Coyne, Nelson, Graham-Kevan, Keister, & Grant, 2010; Gentile, 2012), have made efforts to ''dial back'' their language on this issue and reach out to their colleagues on the opposite side of the debate. We are not so much concerned that some scholars argue violent digital games might increase aggression. Differing opinions could be part of a lively and stimulating debate! Our concern is that the ''harm'' position has, too often, been stated in a way that the current evidence does not yield and is greatly misleading to both the scientific community and general public.

There are myriad reasons why this occurred. Media experience cycles of ''moral panic'' (Ferguson, 2010; Kutner & Olson, 2008) in which they are blamed for all manner of social ills. These panics usually take a familiar pattern with elder adults less inclined to use the new media (including politicians and scholars) making extreme claims of the harmfulness of the new media that is primarily used by youth. As those youth age and become active members of society, the panic dies away, although this can take decades. The professional organizations, particularly the American Psychological Association (APA, 2005) and American Academy of Pediatrics (AAP, 2009), arguably failed particularly in ensuring that objective science was upheld rather than indulging in convenient but hysterical political rhetoric. Policy statements on media by the APA and AAP have been found to be riddled with egregious mistakes, such as inflating the number of studies by a factor of 10 (Ferguson, 2009; Freedman, 2002), failing to cite numerous studies than conflicted with the ''harm'' view (Ferguson, 2010; Hall et al., 2011), and repeating debunked ''scientific urban legend'' claims such as comparisons with smoking and lung cancer (Ferguson, 2010). Arguably, policy statements by the APA and AAP violate their own ethical codes regarding careful and objective dissemination of research-related results to the general public. When drafting their policy statements, both the AAP and APA relied on a narrow group of scholars ideologically invested in the ''harm'' view of media effects. These scholars then often refer back to these policy statements they themselves drafted as a kind of ''echo attribution'' (Rosen & Davison, 2001) to imply an independent review of their work that, in fact, never occurred. The resulting policy statements are noncredible and present a glaring example of the breakdown of the scientific process. They, further, are now directly opposed by independent reviews of scholars not involved in either side of the debate, such as the justices in Brown v. EMA (2011), and by the governments of Australia and Sweden. Just as testimony regarding the ''harmfulness'' of comic books given to governments by mental health professionals in the 1950s now looks to be an example of nannying excess on the part of the scientific community, so too will the existing policy statements of the AAP and APA do little other than to damage the credibility of the field (Hall et al., 2011). We recommend that such policy statements be repealed, and more careful peer-review of policy statements implemented in the future. In fairness, the APA appears to have evidenced some development on this issue, declining to participate in the Brown v. EMA (2011) case and citing inconsistencies in the literature (see Azar, 2010).

Improving Our Methods

While, as we have pointed out earlier, the propagation of extreme statements not supported by the available evidence is a problem of ideological convictions, the key condition enabling this current state of affairs are the insufficient or ambiguous methods employed to measure human aggression (Ritter & Eslea, 2005) or the artificial situations under which games are studied (Williams, 2005). With a corpus of precise and valid measurements for the different aspects of aggressiveness (thoughts, emotions, and behaviors), study results could no longer be subjected to interpretations from drastically different perspectives. We feel that the point of empirical evidence, being to provide definite answers to debatable questions, suffers greatly from the arbitrary methodology and consequentially drawn conclusions in the case of game violence research. We recommend scholars to adhere to two steps: One, not to generalize important findings further than the employed methods would allow (e.g., to consider aggression-related semantic activations simply as associations and not as "aggressive thoughts"). Two, at the same time, to overcome these limitations by developing standards (to ensure objectivity) and focusing research on proper validation of key measurements. We acknowledge there are current attempts to realize this, for example for the Hot Sauce Paradigm (Beier & Kutzner, 2012), and we encourage further investments in these directions.

Conclusions

Media moral panics tend, ultimately, to burn down. This happens, generally, for several reasons. First, as noted, the youth who are used to the new media eventually become the influential elder adults. Being comfortable with the new media, they are less inclined to disparage it or identify it as a source of societal ills (although they may simply replace their new media with their children's new media in a kind of "Goldilocks Effect"). Second, the implication that the new media is a public health crisis crumbles when it becomes plainly apparent that no public health crisis emerged. We have now clearly reached that state, given that the current generation of youth is the least violent or suicidal, and most civically engaged on record, while remaining academically successful (Ferguson, 2010).

We are, thus, most concerned about the academic culture which emerged in the decade of the 2000s in which scholars appeared to be encouraged to make more and more extreme statements about violent digital games and the state of research. We do not believe these statements serve scholars well, and certainly do damage to the field. This does not mean that scholars cannot make arguments that digital games may lead to aggression. Rather, it is a matter that such arguments must be careful, take care not to be alarmist, and ethically note opposing research. We are pleased to see that some scholars are responsibly taking such steps (e.g., Coyne et al., 2010; Gentile, 2012), and look forward to debating them in the future! Historically, research has been focused on a social-cognitive perspective. We believe the field of media effects research could prosper through the adoption of different perspectives, and consideration of specific biological, developmental, and environmental risk and resilience factors.

We conclude by encouraging the field to turn a corner. We advocate a critical debate in which claims about effects of violence in digital games are made (and revoked!) based *only* on existing scientific evidence. We encourage scholars from all perspectives to actively participate, to reach a responsible dialog and constructive debate that could continue to be enriching and invigorating. Transitioning from rigid ideology to something that is perhaps less conclusive but more sophisticated will do much to restore the credibility of this field.

Acknowledgments

The research leading to these results received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 240864 (SOFOGA).

References

Adachi, P. J. C., & Willoughby, T. (2011a). The effect of violent video games on aggression: Is it more than just the violence?. *Aggression and Violent Behavior, 16*, 55–62. doi: 10.1016/j.avb.2010.12.002

Adachi, P. J. C., & Willoughby, T. (2011b). The effect of video game competition and violence on aggressive behavior: Which characteristic has the greatest influence? *Psychology of Violence, 1*, 259–274. doi: 10.1037/a0024908

Adachi, P. J. C., & Willoughby, T. (2012). Do video games promote positive youth development?. *Journal of Adolescent Research, 28*, 155–165. doi: 10.1177/0743558412464522

American Academy of Pediatrics. (2009). Policy statement – media violence. *Pediatrics*, *124*, 1495–1503. doi: 10.1542/peds.2009-2146

American Psychological Association. (2005). *Resolution on violence in video games and interactive media.* Retrieved from http://www.apa.org/about/policy/interactive-media.pdf

Anderson, C. A., Berkowitz, L., Donnerstein, E., Huesmann, L. R., Johnson, J. D., Linz, D., . . . , Wartella, E. A. (2003). The influence of media violence on youth. *Psychological Science in the Public Interest, 4*, 81–110. doi: 10.1111/j.1529-1006.2003.pspi_1433.x

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27–51. doi: 10.1146/annurev.psych.53.100901.135231

Anderson, C. A., & Carnagey, N. L. (2009). Causal effects of violent sports video games on aggression: Is it competitiveness or violent content? *Journal of Experimental Social Psychology, 45*, 731–739. doi: 10.1016/j.jesp.2009.04.019

Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin, A. J., Eubanks, J., & Valentine, J. C. (2004). Violent video games: Specific effects of violent content on aggressive thoughts and behavior. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 36, pp. 199–249). New York, NY: Elsevier. doi: 10.1016/S0065-2601(04)36004-1

Anderson, C. A., Deuser, W. E., & DeNeve, K. M. (1995). Hot temperatures, hostile affect, hostile cognition, and arousal: Tests of a general model of affective aggression. *Personality and Social Psychology Bulletin, 21*, 434–448. doi: 10.1177/0146167295215002

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78*, 772–790. doi: 10.1037/0022-3514.78.4.772

Anderson, C. A., Sakamoto, A., Gentile, D. A., Ihori, N., Shibuya, A., Yukawa, S., ... Kobayashi, K. (2008). Longitudinal effects of violent video games on aggression in Japan and the United States. *Pediatrics, 122*, e1067–e1072. doi: 10.1542/peds.2008-1425.

Anderson, C A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., ... Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin, 136*, 151–173. doi: 10.1037/a0018251.

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*, 146–158. doi: 10.1002/ab.20111

Arriaga, P., Monteiro, M. B., & Esteves, F. (2011). Effects of playing violent computer games on emotional desensitization and aggressive behavior. *Journal of Applied Social Psychology, 41*, 1900–1925. doi: 10.1111/j.1559-1816.2011.00791.x

Australian Government Attorney-General's Department. (2010). *Literature review on the impact of playing violent video games on aggression*. Barton, Australia: Author. Retrieved from http://www.ag.gov.au/Documents/LiteratureReview-VVGs and Aggression-November 2010.pdf

Azar, B. (2010). Virtual violence Researchers disagree about whether violent video games increase aggression. *Monitor on Psychology, 41*. Retrieved from http://www.apa.org/monitor/2010/12/virtual-violence.aspx

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, 20*, 203–208. doi: 10.1002/smi.1015

Ballard, M. E., Hamby, R. H., Panee, C. D., & Nivens, E. E. (2006). Repeated exposure to video game play results in decreased blood pressure responding. *Media Psychology, 8*, 323–341. doi: 10.1207/s1532785xmep0804_1

Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication, 28*, 12–29. doi: 10.1111/j.1460-2466.1978.tb01621.x

Barlett, C. P., Branch, O., Rodeheffer, C., & Harris, R. J. (2009). How long do the short-term violent video game effects last? *Aggressive Behavior, 35*, 225–236. doi: 10.1002/ab.20301

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. doi: 10.1016/j.jesp.2007.10.003

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. doi: 10.1002/ab.20279

Bartholow, B. D., & Anderson, C. A. (2002). Effects of violent video games on aggressive behavior: Potential sex differences. *Journal of Experimental Social Psychology, 38*, 283–290. doi: 10.1006/jesp.2001.1502

Bartholow, B. D., Bushman, B. J., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology, 42*, 532–539. doi: 10.1016/j.jesp.2005.08.006

Bartholow, B. D., Sestir, M. A., & Davis, E. B. (2005). Correlates and consequences of exposure to video game violence: Hostile personality, empathy, and aggressive behavior. *Personality and Social Psychology Bulletin, 31*, 1573–1586. doi: 10.1177/0146167205277205

Beier, R., & Kutzner, F. (2012). *Choose a juice! The effect of choice options and intention on aggression in a modified hot-sauce paradigm.* Paper presented at the 13th Annual Meeting of Society for Personality and Social Psychology, San Diego, CA.

Bennerstedt, U., Ivarsson, J., & Linderoth, J. (2012). How gamers manage aggression: Situating skills in collaborative computer games. *International Journal of Computer-Supported Collaborative Learning, 7*, 43–61. doi: 10.1007/s11412-011-9136-6

Bluemke, M., Friedrich, M., & Zumbach, J. (2010). The influence of violent and nonviolent computer games on implicit measures of aggressiveness. *Aggressive Behavior, 36*, 1–13. doi: 10.1002/ab.20329

Brady, S. S., & Matthews, K. A. (2006). Effects of media violence on health-related outcomes among young men. *Archives of Pediatrics & Adolescent Medicine, 160*, 341–347. doi: 10.1001/archpedi.160.4.341

Breuer, J., Elson, M., Mohseni, M. R., & Scharkow, M. (2012). *Are we really only measuring media effects? Problems and pitfalls associated with the implementation and analysis of the Competitive Reaction Time Task (CRTT) in research on digital games.* Paper presented at the 17th Workshop Aggression, Luxembourg.

Brown v. Entertainment Merchants Association, 564 U.S. __ , 131 S.Ct. 2729 (2011).

Bushman, B. J., & Anderson, C. A. (2011). Weighing the evidence: Comparison of two amicus briefs submitted to U.S. Supreme Court violent video game case. Retrieved from http://www.psychology.iastate.edu/faculty/caa/abstracts/2010-2014/11BA.pdf

Carnagey, N. L., & Anderson, C. A. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition, and behavior. *Psychological Science, 16*, 882–889. doi: 10.1111/j.1467-9280.2005.01632.x

Cicchirillo, V., & Chory-Assad, R. M. (2005). Effects of affective orientation and video game play on aggressive thoughts and behaviors. *Journal of Broadcasting & Electronic Media, 49*, 435–449. doi: 10.1207/s15506878jobem4904_5

Coyne, S. M., Nelson, D. A., Graham-Kevan, N., Keister, E., & Grant, D. M. (2010). Mean on the screen: Psychopathy, relationship aggression, and aggression in the media. *Personality and Individual Differences, 48*, 288–293. doi: 10.1016/j.paid.2009.10.018

Denzler, M., Häfner, M., & Förster, J. (2011). He just wants to play: How goals determine the influence of violent computer games on aggression. *Personality and Social Psychology Bulletin, 37*, 1644–1654. doi: 10.1177/0146167211421176

Eastin, M. S. (2006). Video game violence and the female game player: Self- and opponent gender effects on presence and aggressive thoughts. *Human Communication Research, 32*, 351–372. doi: 10.1111/j.1468-2958.2006.00279.x

Eastin, M. S. (2007). The influence of competitive and cooperative group game play on state hostility. *Human Communication Research, 33*, 450–466. doi: 10.1111/j.1468-2958.2007.00307.x

Elson, M., Breuer, J., Van Looy, J., & Kneer, J. (2012). *Comparing apples and oranges? The effects of confounding factors in experimental research on digital games and aggression*. Paper presented at the 62nd Conference of the International Communication Association, Phoenix, AZ.

Ewoldsen, D. R., Eno, C. A., Okdie, B. M., Velez, J. A, Guadagno, R. E., & DeCoster, J. (2012). Effect of playing violent video games cooperatively or competitively on subsequent cooperative behavior. *Cyberpsychology, Behavior, and Social Networking, 15*, 277–280. doi: 10.1089/cyber.2011.0308

Farrar, K. M., Krcmar, M., & Nowak, K. L. (2006). Contextual features of violent video games, mental models, and aggression. *Journal of Communication, 56*, 387–405. doi: 10.1111/j.1460-2466.2006.00025.x

Ferguson, C. J. (2009). Media violence effects: Confirmed truth or just another X-file?. *Journal of Forensic Psychology Practice, 9*, 103–126. doi: 10.1080/15228930802572059

Ferguson, C. J. (2010). Media violence effects and violent crime: Good science or moral panic? In C. J. Ferguson (Ed.), *Violent crime: Clinical and social implications* (pp. 37–56). Thousand Oaks, CA: Sage.

Ferguson, C. J. (2011). Video games and youth violence: A prospective analysis in adolescents. *Journal of Youth and Adolescence, 40*, 377–391. doi: 10.1007/s10964-010-9610-x

Ferguson, C. J., & Dyck, D. (2012). Paradigm change in aggression research: The time has come to retire the General Aggression Model. *Aggression and Violent Behavior, 17*, 220–228. doi: 10.1016/j.avb.2012.02.007

Ferguson, C. J., & Garza, A. (2011). Call of (civic) duty: Action games and civic behavior in a large sample of youth. *Computers in Human Behavior, 27*, 770–775. doi: 10.1016/j.chb.2010.10.026

Ferguson, C. J., Garza, A., Jerabeck, J., Ramos, R., & Galindo, M. (2013). Not worth the fuss after all? Cross-sectional and prospective data on violent video game influences on aggression, visuospatial cognition and mathematics ability in a sample of youth. *Journal of Youth and Adolescence, 42*, 109–122. doi: 10.1007/s10964-012-9803-6

Ferguson, C. J., & Kilburn, J. (2009). The public health risks of media violence: A meta-analytic review. *The Journal of Pediatrics, 154*, 759–763. doi: 10.1016/j.jpeds.2008.11.033

Ferguson, C. J., & Kilburn, J. (2010). Much ado about nothing: The misestimation and overinterpretation of violent video game effects in Eastern and Western nations: Comment on Anderson et al. (2010). *Psychological Bulletin, 136*, 174–178. doi: 10.1037/a0018566

Ferguson, C. J., & Rueda, S. M. (2009). Examining the validity of the modified Taylor competitive reaction time test of aggression. *Journal of Experimental Criminology, 5*, 121–137. doi: 10.1007/s11292-009-9069-5

Ferguson, C. J., & Rueda, S. M. (2010). The Hitman study: Violent video game exposure effects on aggressive behavior, hostile feelings, and depression. *European Psychologist, 15*, 99–108. doi: 10.1027/1016-9040/a000010

Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S., & Smith, S. M. (2008). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation?. *Criminal Justice and Behavior, 35*, 311–332. doi: 10.1177/0093854807311719

Ferguson, C. J., San Miguel, C., Garza, A., & Jerabeck, J. M. (2012). A longitudinal test of video game violence influences on dating and aggression: a 3-year longitudinal study of adolescents. *Journal of Psychiatric Research, 46*, 141–146. doi: 10.1016/j.jpsychires.2011.10.014

Ferguson, C. J., San Miguel, C., & Hartley, R. D. (2009). A multivariate analysis of youth violence and aggression: The influence of family, peers, depression, and media violence. *The Journal of Pediatrics, 155*, 904–908. doi: 10.1016/j.jpeds.2009.06.021

Ferguson, C. J., & Savage, J. (2012). Have recent studies addressed methodological issues raised by five decades of television violence research? A critical review. *Aggression and Violent Behavior, 17*, 129–139. doi: 10.1016/j.avb.2011.11.001

Ferguson, C. J., Smith, S. M., Miller-Stratton, H., Fritz, S., & Heinrich, E. (2008). Aggression in the laboratory: Problems with the validity of the modified Taylor competitive reaction time test as a measure of aggression in media violence studies. *Journal of Aggression, Maltreatment & Trauma, 17*, 118–132. doi: 10.1080/10926770802250678

Fikkers, K. M., Valkenburg, P., & Vossen, H. G. M. (2012). *Validity of adolescents' direct estimates of their exposure to media violence in three types of media.* Paper presented at the 62nd Conference of the International Communication Association, Phoenix, AZ.

Fischer, J., Aydin, N., Kastenmüller, A., Frey, D., & Fischer, P. (2012). The delinquent media effect: Delinquency-reinforcing video games increase players attitudinal and behavioral inclination toward delinquent behavior. *Psychology of Popular Media Culture, 1*, 201–205. doi: 10.1037/a0028114

Fischer, P., Kastenmüller, A., & Greitemeyer, T. (2010). Media violence and the self: The impact of personalized gaming characters in aggressive video games on aggressive behavior. *Journal of Experimental Social Psychology, 46*, 192–195. doi: 10.1016/j.jesp.2009.06.010

Freedman, J. L. (2002). *Media violence and its effect on aggression: Assessing the scientific evidence*. Toronto, ON: University of Toronto Press.

Frindte, W., & Obwexer, I. (2003). Ego-Shooter: Gewalthaltige Computerspiele und aggressive Neigungen [Violent video games and aggressive tendencies]. *Zeitschrift für Medienpsychologie, 15*, 140–148. doi: 10.1026/1617-6383.15.4.140

Gentile, D. A. (2012). Yet one more look: Rethinking our thinking. *Amplifier, Summer/Spr*, 13–14. Retrieved from http://www.apa.org/divisions/div46/newsletter.html.

Giumetti, G. W., & Markey, P. M. (2007). Violent video games and anger as predictors of aggression. *Journal of Research in Personality, 41*, 1234–1243. doi: 10.1016/j.jrp.2007.02.005

Greitemeyer, T., & McLatchie, N. (2011). Denying humanness to others: A newly discovered mechanism by which violent video games increase aggressive behavior. *Psychological Science, 22*, 659–665. doi: 10.1177/0956797611403320

Greitemeyer, T., Traut-Mattausch, E., & Osswald, S. (2012). How to ameliorate negative effects of violent video games on cooperation: Play it cooperatively in a team. *Computers in Human Behavior, 28*, 1465–1470. doi: 10.1016/j.chb.2012.03.009

Grimes, T., Anderson, J. A., & Bergen, L. A. (2008). *Media violence and aggression. Science and ideology*. Thousand Oaks, CA: Sage.

Gunter, W. D., & Daly, K. (2012). Causal or spurious: Using propensity score matching to detangle the relationship between violent video games and violent behavior. *Computers in Human Behavior, 28*, 1348–1355. doi: 10.1016/j.chb.2012.02.020

Hall, R. C. W., Day, T., & Hall, R. C. W. (2011). A plea for caution: Violent video games, the Supreme Court, and the role of science. *Mayo Clinic Proceedings, 86*, 315–321. doi: 10.4065/mcp.2010.0762

Hall, R. C. W., & Hall, R. C. W. (2011). A further plea for caution against medical professionals overstating video game violence effects. *Mayo Clinic Proceedings, 86*, 821–823. doi: 10.4065/mcp.2011.0357

Happ, C, Melzer, A., & Steffgen, G. (2011). Bringing empathy into play: On the effects of empathy in violent and nonviolent video games. In J. Anacleto, S. Fels, N. Graham, B. Kapralos, M. S. El-Nasr, & K. Stanley (Eds.), *Lecture Notes in Computer Science: Vol. 6972. Entertainment Computing*, (pp. 371–374). doi: 10.1007/978-3-642-24500-8_44

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology, 13*, 339–363. doi: 10.1080/15213269.2010.524912

Hartmann, T., & Vorderer, P. (2010). It's okay to shoot a character: Moral disengagement in violent video games. *Journal of Communication, 60*, 94–119. doi: 10.1111/j.1460-2466.2009.01459.x

Hasan, Y., Bègue, L., & Bushman, B. J. (2012). Viewing the world through "blood-red tinted glasses": The hostile expectation bias mediates the link between violent video game exposure and aggression. *Journal of Experimental Social Psychology, 48*, 953–956. doi: 10.1016/j.jesp.2011.12.019

Hopf, W. H., Huber, G. L., & Weiß, R. H. (2008). Media violence and youth violence: A 2-year longitudinal study. *Journal of Media Psychology: Theories, Methods, and Applications, 20*, 79–96. doi: 10.1027/1864-1105.20.3.79

Huesmann, L. R. (2010). Nailing the coffin shut on doubts that violent video games stimulate aggression: Comment on Anderson et al. (2010). *Psychological Bulletin, 136*, 179–181. doi: 10.1037/a0018567

iLab.o. (2011). *Digimeter, rapport 3. Mediatechnologie en ICT-gebruik in Vlaanderen. Wave 3* [Digimeter, report 3. Media technology and ICT-usage in Flanders. Wave 3]. Ghent, Belgium: IBBT.

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. doi: 10.1111/j.1460-2466.2007.00356.x

Jeong, E. J., Biocca, F. A., & Bohil, C. J. (2012). Sensory realism and mediated aggression in video games. *Computers in Human Behavior, 28*, 1840–1848. doi: 10.1016/j.chb.2012.05.002

Jerabeck, J. M., & Ferguson, C. J. (2012). *Violent video games, aggressive behavior and prosocial behavior.* Unpublished master's thesis, Texas A&M International University, Laredo, TX, USA.

Karvinen, J., & Mäyrä, F. (2011). *Pelaajabarometri 2011: Pelaamisen muutos* [Player barometer 2011: Change in gaming]. Tampere, Finland: University of Tampere, TRIM.

Kneer, J., Glock, S., Beskes, S., & Bente, G. (2012). Are digital games perceived as fun or danger? Supporting and suppressing different game-related concepts. *Cyberpsychology, Behavior, and Social Networking, 15*, 604–609. doi: 10.1089/cyber.2012.0171

Kneer, J., Munko, D., Glock, S., & Bente, G. (2012). Defending the doomed: Implicit strategies concerning protection of first-person shooter games. *Cyberpsychology, Behavior, and Social Networking, 15*, 251–256. doi: 10.1089/cyber.2011.0583

Koglin, U., Witthöft, J., & Petermann, F. (2009). Gewalthaltige Computerspiele und aggressives Verhalten im Jugendalter [Violent computer games and aggressive behavior in adolescence]. *Psychologische Rundschau, 60*, 163–172. doi: 10.1026/0033-3042.60.3.163

Konijn, E. A., Nije Bijvank, M., & Bushman, B. J. (2007). I wish I were a warrior: The role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Developmental Psychology, 43*, 1038–1044. doi: 10.1037/0012-1649.43.4.1038

Krahé, B., Möller, I., Huesmann, L. R., Kirwil, L., Felber, J., & Berger, A. (2011). Desensitization to media violence: Links with habitual media violence exposure, aggressive cognitions, and aggressive behavior. *Journal of Personality and Social Psychology, 100*, 630–646. doi: 10.1037/a0021711

Kutner, L. A., & Olson, C. K. (2008). *Grand theft childhood. The surprising truth about violent video games and what parents can do*. New York, NY: Simon & Schuster.

Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior, 25*, 331–348. doi: 10.1002/(SICI)1098-2337(1999)25:5<331::AID-AB2>3.0.CO;2-1

Möller, I., & Krahé, B. (2009). Exposure to violent video games and aggression in German adolescents: A longitudinal analysis. *Aggressive Behavior, 35*, 75–89. doi: 10.1002/ab.20290

Pinker, S. (2002). *The blank slate. The modern denial of human nature*. New York, NY: Penguin Books.

Puri, K., & Pugliese, R. (2012). Sex, lies, and video games: Moral panics or uses and gratifications. *Bulletin of Science, Technology & Society, 32*, 345–352. doi: 10.1177/0270467612463799

Quandt, T., Scharkow, M., & Festl, R. (2010). Digitales Spielen als mediale Unterhaltung [Digital games as media entertainment]. *MediaPerspektiven, 11*, 515–522.

Ravaja, N., & Kivikangas, M. J. (2009). Designing game research: Addressing questions of validity. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (pp. 404–412). New York, NY: Routledge.

Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior, 31*, 407–419. doi: 10.1002/ab.20066

Rosen, G. M., & Davison, G. C. (2001). "Echo attributions" and other risks when publishing on novel therapies without peer review. *Journal of Clinical Psychology, 57*, 1245–1250. doi: 10.1002/jclp.1092

Rothmund, T., Gollwitzer, M., & Klimmt, C. (2011). Of virtual victims and victimized virtues: Differential effects of experienced aggression in video games on social cooperation. *Personality and Social Psychology Bulletin, 37*, 107–119. doi: 10.1177/0146167210391103

Sacks, D. P., Bushman, B. J., & Anderson, C. A. (2011). Do violent video games harm children? Comparing the scientific amicus curiae "experts" in Brown v. Entertainment Merchants Association. *Northwestern University Law Review Colloquy, 106*, 1–12.

Saleem, M., Anderson, C. A., & Gentile, D. A. (2012). Effects of prosocial, neutral, and violent video games on college students' affect. *Aggressive Behavior, 38*, 263–271. doi: 10.1002/ab.21427

Savage, J. (2004). Does viewing violent media really cause criminal violence? A methodological review. *Aggression and Violent Behavior, 10*, 99–128. doi: 10.1016/j.avb.2003.10.001

Schmierbach, M. (2010). "Killing spree": Exploring the connection between competitive game play and aggressive cognition. *Communication Research, 37*, 256–274. doi: 10.1177/0093650209356394

Sestir, M. A., & Bartholow, B. D. (2010). Violent and nonviolent video games produce opposing effects on aggressive and prosocial outcomes. *Journal of Experimental Social Psychology, 46*, 934–942. doi: 10.1016/j.jesp.2010.06.005

Sherry, J. L. (2001). The effects of violent video games on aggression: A meta analysis. *Human Communication Research, 27*, 409–431. doi: 10.1111/j.1468-2958.2001.tb00787.x

Sherry, J. L. (2007). Violent video games and aggression: Why can't we find effects? In R. Preiss, B. Gayle, N. Burrell, M. Allen, & J. Bryant (Eds.), *Mass media effects research: Advances through meta-analysis* (pp. 245–262). Mahwah, NJ: Erlbaum.

Slater, M. D., Henry, K. L., Swaim, R. C., & Anderson, L. L. (2003). Violent media content and aggressiveness in adolescents: A downward spiral model. *Communication Research, 30*, 713–736. doi: 10.1177/0093650203258281

Statens Medieråd. (2011). *Våldsamma datorspel och aggression. En översikt av forskningen 2000–2011* [Violent computer games and agggression. An overview of the research 2000–2011]. Stockholm, Sweden: Statens Medieråd. Retrieved from http://www.statensmedierad.se/upload/_pdf/Statens_medierad_Valdsamma_datorspel_och_aggression.pdf

Strasburger, V. C. (2007). Go ahead punk, make my day: It's time for pediatricians to take action against media violence. *Pediatrics, 119*, e1398–e1399. doi: 10.1542/peds.2007-0083

Surette, R. (2012). Cause or catalyst: The interaction of real world and media crime models. In *American Journal of Criminal Justice*. Advance online publication. doi: 10.1007/s12103-012-9177-z

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality, 35*, 297–310. doi: 10.1111/j.1467-6494.1967.tb01430.x

Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior, 1*, 163–177. doi: 10.1016/1359-1789(95)00014-3

Uhlmann, E., & Swanson, J. (2004). Exposure to violent video games increases automatic aggressiveness. *Journal of Adolescence, 27*, 41–52. doi: 10.1016/j.adolescence.2003.10.004

Unsworth, G., Devilly, G. J., & Ward, T. (2007). The effect of playing violent video games on adolescents: Should parents be quaking in their boots? *Psychology Crime & Law, 13*, 383–394. doi: 10.1080/10683160601060655

Valadez, J. J., & Ferguson, C. J. (2012). Just a game after all: Violent video game exposure and time spent playing effects on hostile feelings, depression, and visuospatial cognition. *Computers in Human Behavior, 28*, 608–616. doi: 10.1016/j.chb.2011.11.006

Velez, J. A., Mahood, C., Ewoldsen, D. R., & Moyer-Guse, E. (2012). Ingroup versus outgroup conflict in the context of violent video game play: The effect of cooperation on increased helping and decreased aggression. In *Communication Research*. doi: 10.1177/0093650212456202

Von Salisch, M., Vogelgesang, J., Kristen, A., & Oppl, C. (2011). Preference for violent electronic games and aggressive behavior among children: The beginning of the downward spiral? *Media Psychology, 14*, 233–258. doi: 10.1080/15213269.2011.596468

Wallenius, M., & Punamäki, R.-L. (2008). Digital game violence and direct aggression in adolescence: A longitudinal study of the roles of sex, age, and parent-child communication. *Journal of Applied Developmental Psychology, 29*, 286–294. doi: 10.1016/j.appdev.2008.04.010

Ward, M. R. (2011). Video games and crime. *Contemporary Economic Policy, 29*, 261–273. doi: 10.1111/j.1465-7287.2010.00216.x

Williams, D. (2005). Bridging the methodological divide in game research. *Simulation & Gaming, 36*, 447–463. doi: 10.1177/1046878105282275

Willoughby, T., Adachi, P. J. C., & Good, M. (2012). A longitudinal study of the association between violent video game play and aggression among adolescents. *Developmental Psychology, 48*, 1044–1057. doi: 10.1037/a0026046

Zhen, S., Xie, H., Zhang, W., Wang, S., & Li, D. (2011). Exposure to violent computer games and Chinese adolescents' physical aggression: The role of beliefs about aggression, hostile expectations, and empathy. *Computers in Human Behavior, 27*, 1675–1687. doi: 10.1016/j.chb.2011.02.006

Received August 8, 2012
Accepted February 14, 2013
Published online December 10, 2013

About the authors

Malte Elson is a psychologist and works as a research associate in the European Research Council (ERC) funded project "Social Foundations of Online Gaming" at the Department of Communication, University of Münster, Germany. His research interests include the effects of digital games, media content regulation, and behavior research methods.

Christopher J. Ferguson is department chair of psychology at Stetson University, DeLand, FL, USA. He has published extensively on the topic of video game violence and was invited to comment on the topic at US Vice President Biden's gun violence hearings following the tragic Sandy Hook shooting incident of December 14, 2012.

Malte Elson

Department of Communication
University of Münster
Bispinghof 9–14
48143 Münster
Germany
Tel. +49 251 83-21263
Fax +49 251 83-21310
E-mail malte.elson@uni-muenster.de

Does Doing Media Violence Research Make One Aggressive?

The Ideological Rigidity of Social-Cognitive Theories of Media Violence and a Response to Bushman and Huesmann (2013), Krahé (2013), and Warburton (2013)

Malte Elson[1] and Christopher J. Ferguson[2]

[1]Department of Communication, University of Münster, Germany, [2]Department of Psychology, Stetson University, DeLand, FL, USA

We thank Bushman and Huesmann (2013), Krahé (2013), and Warburton (2013) for taking the time to comment on our review of digital game violence research. This is obviously a field where considerable controversy continues to exist and an opportunity for cordial debate could help resolve differences in the field. The current crop of comments run the gamut from keeping to reasonable points of disagreement (Warburton), to relying on sometimes snide comments ("The art of omission," Krahé), to the comment by Bushman and Huesmann which makes use of ad hominem attacks. They also tend to restate old arguments that have been discredited either in our review or other past reviews (e.g., Adachi & Willoughby, 2011; Hall, Day, & Hall, 2011a, 2011b). Some of these statements confirm our initial concerns that the debate on violent media effects has shifted from science to ideology.

In the following, we examine some of the statements by our colleagues that we find to be problematic or misleading.

Theoretical Perspectives on the Effects of Violent Games

Bushman and Huesmann start out by explaining that there is abundant evidence for the harmful effects of observing violence in the home, at school, in the community, or in the culture on children. They refer to studies investigating the development of aggression through observation of violence in the Israeli-Palestinian conflict (Boxer et al., 2013), or the effects of exposure to rocket attacks on distress and violence (Henrich & Shahar, 2013). Then they turn to the audience by asking a rhetorical question (p. 3): "How, then, could viewing violence in the mass media not be harmful to children?" They ask for a psychological theory that explains how the risk of violence is increased by observing violence in the home, school, community, or culture, but not by observing it in the media.

We, in turn, might ask what one thing has to do with the other. Bushman and Huesmann imply two things here: (1) The experience or observation of real acts of violence is qualitatively similar to those of fictional violence. This is quite a stretch given that children aged five or younger are already able to distinguish between real and fictional television (Wright, Huston, Reitz, & Piemyat, 1994).

(2) Observing proximal acts of violence (e.g., in the family) has similar psychological effects as watching virtual displays of violence (e.g., in a digital game). This argument is certainly not supported by the clinical practice of this comment's second author, in which he has regularly observed the devastating influence of family violence on children, but cannot think of a single case in which watching Woody Woodpecker or playing Call of Duty was the root of a child's mental health suffering. This argument can be dismissed by considering a simple, but obviously unethical and illegal hypothetical experiment. Take 200 children and randomize 100 to watch their parents viciously

DOI: 10.1027/1016-9040/a000185

attack one another for a hour a day, the other 100 to watch a violent television program an hour a day, then assess their mental health after one month is over. Although this experiment will obviously always remain in the realm of the hypothetical (one hopes at least), to suggest the mental health outcomes for these children would be even remotely identical is absurd.

Bushman and Huesmann then respond to our review of explanatory models and theories. Apparently, their biggest criticism of the Catalyst Model that we discuss as one theoretical perspective is that it is not ''new.'' This assertion is correct, given that it was published 5 years ago (Ferguson, Rueda, et al., 2008) and is certainly supported by recent empirical investigations (e.g., Ferguson, Ivory, & Beaver, 2013; Surette, 2013). Perhaps they mean that, as a diathesis-stress model, it is similar to other diathesis-stress models, which is certainly the case.

Bushman and Huesmann and Krahé accuse us of selectively citing only evidence for the Catalyst Model (the ''selection path'') and against an effect of violent games on aggression (the ''socialization path''). This claim is simply nonfactual. One might take a look at the reference list of our review article to find well over 30 publications in which research questions and results are explicitly explained with the General Aggression Model (GAM; Anderson & Bushman, 2002). Conversely, neither of them cites any existing evidence or perspectives opposing their own views in their comments (other than a select few to criticize). As one example, readers might compare Dr. Bushman's section of a recent National Science Foundation (NSF) report on gun violence (Subcommittee of Youth Violence, 2013) to a review of media violence almost simultaneously released by Common Sense Media (CSM, 2013), a traditionally anti-media watchdog group. Although we do not agree with the conclusions of CSM, we admire the latter as a model for making an honest and balanced argument for the ''harm'' perspective. We invite readers to consider these two contemporary reports side-by-side, and examine their differences.

They reject our criticism of the GAM mainly on three grounds: First, that the GAM is a social-cognitive model intended to be used by aggression researchers, not clinicians in the area of pathological aggression. This is in stark contrast with the degree to which these authors themselves generalize the GAM to criminal behavior (Huesmann, 2013) and clinical realms (see, once again, Subcommittee of Youth Violence, 2013). Bushman and Huesmann underline the historical importance of the GAM for the field of aggression research. We are certainly well aware of its impact (although we do note it is seldom used in fields outside of media effects such as criminology), which is why we presented it as one theoretical approach to media violence effects research in our literature review. However, we are uncertain where exactly (and why) the GAM draws the line between the general kind of aggression that researchers are supposedly interested in, and the pathological aggressiveness only clinicians are concerned with. It also does not reflect recent developments in clinical psychology and diagnostics (such as the DSM-5; American Psychiatric Association, 2013), which tend to understand more and more disorders as a spectrum, and not as distinctively ''pathological'' or ''non-pathological.'' Proponents of the GAM should explicate the subgroups of the population and the kinds of behaviors covered by the GAM, and those ''too pathological.'' We believe, however, that this could severely limit the significance and relevance of the GAM, as the general public is concerned with antisocial, violent, or criminal behaviors, rather than higher-order cognitive processes.

We endorse Warburton's call for empirical research that examines the boundaries of media effects, and would like to take this one step further to theoretical and conceptual work. We believe that the strength of every theory or model lies within the differentiated and accurate description of its boundaries and limitations. For the GAM, however, we are currently observing the opposite, as there are attempts to extend it to other areas, such as suicide and global warming (DeWall, Anderson, & Bushman, 2011). Nevertheless, we welcome Warburton's suggestion and see this as a potential starting point for a fruitful debate: What kind of processes can be explained by the GAM, and for which behaviors are other theories or models (such as the Catalyst Model) more suited? Or, indeed, can we advance our understanding of media effects research by reconsidering theoretical models altogether? Is it time to replace social-cognitive theories of media effects such as the GAM with more user-driven media theories such as Uses and Gratifications (Sherry, Lucas, Greenberg, & Lachlan, 2006)?

The second argument is that there are strong effects of social-cognitive processes and priming on behavior, for which Krahé cites the famous study by Bargh, Chen, and Burrows (1996) as one example. While discussing the mechanisms of such cognitive activations is beyond the scope of this response, we point to recent failed direct and conceptual replications of the Bargh et al. study (Doyen, Klein, Pichon, & Cleeremans, 2012) and others (Pashler, Coburn, & Harris, 2012; Shanks et al., 2013), which put the robustness of priming effects on behavior into question. Thus, we see a different similarity between the fields of social priming and media violence than Krahé: Both have been considered ''received wisdom'' and are now experiencing replication crises against which their proponents appear to be extremely defensive. Far from ''easily refuting'' (p. 5) our comments, Krahé's observation reinforces our concerns about the overstatement of mixed research results.

Third, Bushman and Huesmann argue that social learning through observation of violence in the school, home, or community is a strong predictor of aggressive and violent behavior. It is wrong to assume that we are denying the effectiveness of social learning, which is also why, in fact, family and peer violence is an important risk factor in the Catalyst Model. This misconception probably stems from the fact that we consider violent games (and other media) to be rather limited as ''teachers'' or models compared to other influences (e.g., parents). The GAM certainly does not discern between different modes of observation (virtual vs. real) nor different types of observed violence (fictional vs. nonfictional). It fails to explain how observations of fictional violence in virtual worlds (e.g., soldiers fighting) generalize to actual behaviors that are completely different in the real world (e.g., domestic abuse). Also, it equates media experiences with real experiences (e.g., watching violent cartoons vs. watching parents fighting), as Bushman and Huesmann, and Warburton do in their comments. We find this to be a major weakness, not a strength of their model.

The GAM also does not specify the outcome when one observes competing behavioral models in different contexts, for example, rewards for aggressive behavior in games, and punishment for aggressive behavior/rewards for prosocial behavior in the family. The GAM does in no way explain which of the two models would be the "better" or more important one, and why. Even under the assumption violent games were effective models for behavior, children would progressively express aggressiveness in the school or family, for which they would usually get punished, and as such, observe and learn that aggressive acts lead to undesired consequences. If on the other hand antisocial behaviors are tolerated or rewarded, does not the real issue lie with an unhealthy environment that fosters aggressiveness rather than peacefulness?

Empirical Evidence and Methodological Issues

The perspectives on empirical evidence by the commenters can be divided into two sections. The first regards the appropriateness and validity of common methodological procedures in the research. They defend the Competitive Reaction Time Task (CRTT) in particular on the ground that it supposedly has a high *experimental* or *psychological realism*. However, neither of them cites any study supporting this argument for the CRTT. Conversely, they do avoid mentioning the study by Mitchell (2012) which revisits the "truth or triviality" issue and presents a gap between the results of laboratory and field studies in aggression research (and other areas of psychological science). There is no evidence for Warburton's claim that participants actually believe they can *hurt* another with the CRTT, which is certainly a requirement for its validity. We remain skeptical whether this proves true. The CRTT noises are certainly unpleasant, but they are far from being harmful. Moreover, actually harmful noise blasts would severely limit CRTT's further use in laboratory experiments for ethical reasons.

Krahé claims that we quote from one unpublished paper to support our criticism of the CRTT, when in truth we are actually citing at least four publications which discuss its lack of standardization and issues in reliability and validity (Ferguson & Rueda, 2009; Ferguson, Smith, et al., 2008; Savage, 2004; Tedeschi & Quigley, 1996), plus one study showing that there is no link between habitual violent media exposure and CRTT scores (Krahé et al., 2011). The study by Breuer, Elson, Mohseni, and Scharkow (2012) that Krahé refers to supports concerns about CRTT's psychometric objectivity. Instead of providing evidence for the CRTT's external validity, Bushman and Huesmann refer to the example of prisoners being punished with loud music at Guantanamo Bay. To us, however, this torture scenario, involving non-consenting prisoners exposed to hours upon hours of sleep depriving noise, only vaguely resembles the CRTT with its brief exposure and ostensibly consenting opponents in university laboratories. Equating the CRTT to torture seems, to us, one more example of the irresponsible overreach to which this field has become accustomed.

Krahé also criticizes us for avoiding the study by Giancola and Parrott (2008) in support of its validity. In fact, we did not cite this study because it is largely irrelevant to the argument. First, Giancola and Parrott used electrical shocks, not any of the numerous noise-burst variants which are the norm in media effects research. To the best of our knowledge, the electroshock version of the CRTT has never been used in experiments on violent game effects. Instead, media effects researchers usually rely on non-painful noise-bursts instead. We remain skeptical as to whether these two very different stimuli should be equated. Second, the study does in no way provide any evidence for *external* or *construct* validity, but only for its *group discrimination* validity (intoxicated vs. sober participants with a high self-reported propensity toward aggression), which as Tedeschi and Quigley (2000) state is not only weak evidence for validity, but also might easily constitute a logical fallacy. Giancola and Parrott also do not address the issue of lacking standardization, which is certainly an issue for the CRTT's psychometric objectivity. None of the comments cited any evidence to alleviate our concern that there is a lack of validity data for the CRTT. Therefore, the CRTT should not be generalized to significant real-world aggression. Unfortunately, all too often, scholars do exactly this.

Krahé rejects our call for rather using clinical measures of aggression because they would be inappropriate for community samples of children, adolescents, and adults. We are puzzled by this statement given that such measures are, in fact, normed on community samples and perfectly valid for use with all children. The avoidance of measures that actually document harm, and the reliance on unstandardized, non-validated laboratory procedures such as the CRTT, is particularly problematic when, once again, a field appears very willing to generalize its results to clinical outcomes of "harm to children."

Krahé underlines the importance of "experimental research in artificial settings (. . .) with a tighter control of confounding or distorting factors that would affect behavior under natural conditions" (p. 7), and we wholeheartedly agree: Proper experimental laboratory research is key to understanding basic relationships between variables. Consequentially, however, one should only generalize these findings to less artificial and more naturalistic situations with due care. This is something we see repeatedly disregarded in effects research on violent games. This observation also assumes that those confounding factors in game effects experiments have been properly controlled, which has been questioned as well (Adachi & Willoughby, 2011; Elson, Breuer, Van Looy, Kneer, & Quandt, in press).

Warburton's major argument is to put the findings on violence in games into context of other media effects research, with the effects of advertisement on food choice as an example. Warburton states that those comparisons should only be made unless there is "a valid reason why different psychological mechanisms would underlie those effects" (p. 3). Given the title of his comment, "Apples, oranges, and the burden of proof," we must point to the many differences between the psychological mechanisms underlying the role of advertising and fictional media.

Briefly, to work, advertising need only nudge behavior slightly, from one product to another. Advertisement works particularly well when most people are already motivated to purchase the advertised good, such as food, and not in generating desire for something they have no reason to buy in the first place. Choosing one brand over another is a low-impact choice, while choosing to behave aggressively is not. It does not require the sorts of fundamental changes to motivation or personality often suggested as the outcome of media violence. Second, advertising always relates to the everyday life of the consumer, and thus purports to be "true" (although it often is not). Fictional media rarely resemble everyday experiences of their users, and they only seldom attempt to appear non-fictional. Unlike fictional media, advertisement also addresses the consumer directly by promising a "better life" (e.g., faster cars, healthier food). It is quite reasonable to speculate that advertising works hard to circumnavigate "fiction detectors" in ways fictional media does not. Indeed, equating advertising to fictional media involves the problematic assumption that our brains are incapable of distinguishing fictional from non-fictional sources of media. Indeed, the lab of the second author, applying usual skepticism, has found evidence for advertising effects (Ferguson, Muñoz, & Medrano, 2012) but has been unable to find evidence for many fictional media effects. Thus, calls to equate advertising and fictional media should be rejected as too simplistic.

One reason for Warburton's suggestion of an overall consistency in scientific findings on violent media effects might be his misconception of the meaning of p-values in null hypothesis significance testing (NHST). p-Values do certainly not determine whether an effect is "real" or not, as Warburton claims (p. 11). His description implies several common misinterpretations, such as p being the probability that the H_0 is false, and that statistically significant differences are always relevant. NHST and p-values are not a test of "reality"! We clarify: The p-value is the probability of the observed result or more extreme results *under the assumption that the null hypothesis is true*. Given that problems with p-values and NHST (and common misconceptions thereof) have been debated for more than 70 years in psychology (and other sciences), and are beyond the scope of this comment, we recommend Cohen (1994) for a further look.

Meta-Analyses

The second argument regards previously published reviews and meta-analyses of the empirical evidence. Bushman and Huesmann criticize us for not conducting a meta-analysis to test the Catalyst Model, and instead supposedly using an "informal vote-counting" approach. We are not sure where they got that impression. The structure in each section of our empirical review is similar: We first present empirical studies investigating the simple relationship between violence in games and aggressive outcomes (thoughts, emotions, and behaviors). Then, we put these findings into the context of other studies looking at other factors potentially influencing relevant (in-)dependent variables. We are, hence, not "missing the point of moderation" (see Krahé's comment, p. 6), but making it: Researchers must take these third variables (or "distorting factors," see above) into account to avoid issues in validity when studying the effects of violence in games.

We would like to point out several issues with the meta-analysis by Anderson et al. (2010) that Bushman and Huesmann and Warburton discuss in their comments. First, the "average effect size wins!" approach potentially conceals failed replications. Second, the "best practices" defined by Anderson et al. (2010) are rather ambiguous and nontransparent. For example, studies were rated on whether "the outcome measure was appropriate for testing the hypotheses" (p. 9), without providing a clear definition of what can be considered "appropriate" (such as sufficient standardization and validation). Moreover, due to the common use of unstandardized and unvalidated measures (as discussed earlier), a meta-analysis certainly has the same limitations as the studies it includes, an issue Anderson et al. (2010) failed to consider in their "best practices." Also, the authors make much mention of Dr. Rothstein's comments on meta-analysis without specifically noting that she was a coauthor of the Anderson et al. (2010) meta-analysis, and thus not an independent commentator.

Further, and in some ways surprising to us, is the way in which Bushman and Huesmann describe how unpublished material was gathered for the Anderson et al. meta-analysis. It appears that they only asked a limited number of authors for unpublished data supplementary to publications they already identified, and that they did not solicit unpublished studies. In another paper by Rothstein and Bushman (2012), it is explicitly recommended to "include unpublished studies whenever it is possible" (p. 135) and that usually "authors of meta-analyses try to contact every author who has ever published an article on the topic of interest (...)" (p. 131) for unpublished material. We do not wish to speculate why, then, in the case of the Anderson et al. meta-analysis they did not fully heed their own advice. And while it might be true that only 2 out of 88 included papers that reported aggressive behavior were unpublished (although this is not mentioned in the meta-analysis itself), we would like to point out that Appendix A of Anderson et al. (2010) includes at least 16 unpublished studies, and Appendix B 27 unpublished studies.[1] Many of these are non-English technical reports or conference presentations that would be difficult to locate, making replication of the Anderson et al. meta-analysis unlikely. In an earlier comment on their meta-analysis, it was noted that the majority of these unpublished studies came from the authors of the meta-analysis or their collaborators (Ferguson & Kilburn, 2010). At the time, the second author happened to be in touch with the authors of Anderson et al. (2010) on other matters and at no time was asked for unpublished data, nor informed they were conducting a meta-analysis. Since many of these then unpublished data have subsequently been published, their existence and the failure of Anderson et al. to secure them are irrefutable.

We appreciate that Bushman and Huesmann are forthcoming in acknowledging that they made little effort to secure unpublished studies from a wider range of authors in the Anderson et al. meta-analysis, despite Rothstein and Bushman's (2012) advice to do exactly this. This reinforces our concerns about the selection bias against scholarly groups questioning the ''harm'' belief in that study, which may have influenced its conclusions. Even in that meta-analysis, however, effect sizes tend to be truncated with even simple controls (longitudinal effects drop to a meager $r = .075$, for instance, controlling only for gender and time 1 aggression), a fact the authors often fail to note.

No Consensus on a Consensus

Bushman and Huesmann present some preliminary data from a survey, showing that surveyed members of the American Psychological Association's (APA) Media Psychology and Technology Division, the International Communication Association's (ICA) Mass Communication Division, and the American Academy of Pediatrics (AAP) largely agree that violent games increase aggression in children. This is somewhat contrasted by a survey on digital game researchers conducted by the ICA's Game Studies Interest Group, the European Communication Research and Education Association's (ECREA) Digital Games Research Temporary Working Group, and the Digital Games Research Association (DiGRA) (Van Looy et al., 2013). Their sample comprised 544 games researchers, of whom only 1.3% strongly agreed and 8.8% agreed that the effects of digital games on aggressive behavior are a problem for society (27% were undecided, 35.5% disagreed, and 27.6% strongly disagreed). The 64 respondents that indicated psychology as their research tradition did not significantly diverge from that overall results (1.6%; 9.4%; 31.3%; 29.7%; 28.1%; respectively). Thus, the results presented by Bushman and Huesmann were not replicated, and we find support for our initial statement that there is, in fact, disagreement among the research community. This has also been expressed recently in an open letter[2] signed by more than 200 scholars that was sent to the APA, urging its task force on violent media to repeal strong claims made in previous policy statements, and to acknowledge the diverse opinions and perspectives that exist on media violence effects.

Despite our criticism of this kind of rhetoric, Bushman and Huesmann repeat the comparisons between the effect sizes of violent media on aggression and smoking on lung cancer, and justify it by stating that ''calculations don't lie'' (p. 18). Or perhaps they do. The problems with the calculations made to support such conclusions have, by now, been well documented (Block & Crain, 2007; Ferguson, 2009) which they fail to mention. Even ignoring the problematic statistics underlying these comparisons, methodologies of media effects research and oncology are so drastically different that a comparison of the resulting effect sizes is invalid. If cancer studies would consist of participants smoking cigarettes for 5–10 min and then rating their cancer severity on a 5-point Likert scale, then yes, such analogies would be eligible. But fortunately, cancer research does not have the methodology or validity issues that media effects studies do. Ironically, tests for cancer have everything that currently employed aggression tests do not. They are standardized, they are clinically validated (according to the results, one either has cancer or not), and they have a high reliability and external validity (someone who has cancer in a laboratory also has it outside the laboratory). Unfortunately, the same cannot be said about measurements of aggression.

What is, perhaps, most disappointing about the comments of Bushman and Huesmann is that they spend so much time disparaging those who disagree with them. They attempt to resurrect the now-discredited Pollard Sacks, Bushman, and Anderson (2011), despite it had been debunked by scholars uninvested in either side of the debate (Hall, Day, & Hall, 2011b). Put simply, that Bushman and his colleagues should nominate themselves as the ''true experts'' is neither surprising, nor illuminating, and certainly not part of credible science. But let us imagine that their claims of publishing more than skeptics are true (despite Hall et al., 2011b). So what? The fact that at one point in time a certain belief is expressed in a majority of publications is hardly any proof for the validity of this opinion. Looking back at the history of psychology (and other sciences), there are many paradigms that once were particularly popular and influential, and then later regarded as insufficient or simply wrong. One might think of theories such as phrenology or humorism, or to give a more recent (and appropriate) example from psychological science, the cognitive revolution as a response to the once dominant radical behaviorism. All these theories have gone through a paradigm shift in which younger scholars overthrow the ideas of older scholars. We believe that violent media effects research is currently facing the same process.

1 Unpublished study meaning, according to Bushman, Rothstein, and Anderson (2010), ''not published in a peer-reviewed journal, although it could have been published in another outlet.'' (p. 182)

2 See http://www.christopherjferguson.com/APA%20Task%20Force%20Comment1.pdf

It appears to us that Bushman and Huesmann make claims of consensus simply by discounting anyone whom they disagree with. Indeed, toward the end of that piece that appears to include anyone who consumes "large amounts" (undefined, of course) of violent media, or even the authorship of novels (although the second author certainly appreciates the plug by Bushman and Huesmann). We invite readers to consider what a broad brush this is to paint with. They also discount the opinions of the US Supreme Court (Brown v. Entertainment Merchants Association, 2011) and presumably numerous lower courts as well as government reviews by Australia (2010) and Sweden (Statens Medieråd, 2011). Bushman and Huesmann incorrectly imply that scientific evidence had little to do with the US Supreme Court's decision, despite that the majority decision made clear their (rightful) skepticism of the application of this research to a public health issue. They imply the Supreme Court accepted "industry arguments" ignoring that numerous amicus briefs were filed against the "harm" position, by scholars, attorney generals, legal scholars, and youth advocacy groups. Warburton also implies the Australian report was influenced by pressure from the "gaming lobby." Thus, both comments blame any differences in opinion on the gaming community or media industry, instead of arguing what might have been wrong scientifically in these two reports.

It Is Time to Change the Culture of Media Violence Research

We wish to be clear that we are not against scholars making an argument linking media violence with aggression. Our concern remains that the culture of this field has evolved to tolerate sweeping statements equating weak research with public health crises and stifling any form of dissent. Indeed, while functioning as reviewers, we have seen examples of reviews which viciously attack findings which question the "harm" perspective. Several scholars are publically endorsing what is, in effect, censorship of views they disagree with (Gentile, 2013) by demanding that "naysayers" to media violence effects should not be given "valuable (and undeserved)" public attention (Strasburger & Donnerstein, 2013, p. 3).

We find all these observations to be deeply troubling for the credibility of this field. Once again, to be clear, we believe that many scholars on all sides of this debate are doing good work and are dedicated to an open exchange of views, an openness that is at the heart of the scholarly enterprise. But we also observe that some scholars actively and aggressively attempt to quell dissenting views, disparage skeptics, question the motives of those who disagree with them, and enforce a highly ideological view of this field. We believe these efforts have done considerable damage to the scholarly enterprise and the reputation of this field (Hall, Day, & Hall, 2011a). This leads us to wonder what it is about doing aggression research that seems to make some scholars so aggressive. We hope that the majority of scholars will join with us, whatever their personal views may be, in rejecting such a hardline ideological approach to this field and allow it to return to a proper atmosphere of respectful exchange of ideas.

We also express some concern with what appears, to us, to be an overly mechanistic perspective on human learning. This is exemplified by Warburton's comments on human learning, often expressed in language of rigid certitude (i.e., "It is known that…"). We certainly do not deny that humans learn and often learn socially, but we express concern that Warburton's language has converted social learning from something humans can do to something they must do. Warburton also relies on the problematic area of neuroimaging (Vul, Harris, Winkielman, & Pashler, 2009) to support his conclusions in this regard, but overall we find this approach to human learning unsatisfying. To us it is just as important to understand when humans do not learn as when they do, how they make decisions about when to learn, and when to ignore a learning opportunity. Neglecting this in favor of a mechanistic "monkey see/monkey do" model (a metaphor actually used by Orue et al., 2011 to describe their results), to us, does not remotely begin to capture the subtlety and sophistication of human learning and human existence.

More fundamentally, it may be time for this field to consider serious changes in both theory and in communicating to the public. Several prominent media scholars recently headlined a panel entitled "Why don't they believe us?" at the International Communication Association conference (Donnerstein, Strasburger, Viner, & Gentile, 2013). The most parsimonious answer to this question is, in fact, "Because the data are not convincing." Much like psychoanalysts of ages past, media scholars have taken to constructing elaborate theories for why people have not accepted their theories, or even personally attacking those who disagree with them. We contend that the traditional media effects paradigm has failed for the simple reason that it does not comport its own predictions of societal developments. Current theories are arguably too mechanistic, assume viewers are passive receptacles of learning, rather than active shapers and processors of media culture. We do not believe data support the traditional paradigm. We do no less than call on scholars to move past the traditional media effects paradigm, and to an understanding of the interaction between media, behavior, and culture, that is shaped by media users, not media content.

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 240864 (SOFOGA).

References

Adachi, P. J. C., & Willoughby, T. (2011). The effect of violent video games on aggression: Is it more than just the violence? *Aggression and Violent Behavior, 16*, 55–62. doi: 10.1016/j.avb.2010.12.002

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27–51. doi: 10.1146/annurev.psych.53.100901.135231

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., ... Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin, 136*, 151–173. doi: 10.1037/a0018251

Australian Government Attorney-General's Department. (2010). *Literature review on the impact of playing violent video games on aggression*. Barton, Australia: Author. Retrieved from http://www.ag.gov.au

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230–244. doi: 10.1037/0022-3514.71.2.230

Block, J. J., & Crain, B. R. (2007). Omissions and errors in "Media violence and the American public". *American Psychologist, 62*, 252–253. doi: 10.1037/0003-066X.62.3.252

Boxer, P., Huesmann, L. R., Dubow, E. F., Landau, S. F., Gvirsman, S. D., Shikaki, K., & Ginges, J. (2013). Exposure to violence across the social ecosystem and the development of aggression: A test of ecological theory in the Israeli-Palestinian conflict. *Child Development, 84*, 163–177. doi: 10.1111/j.1467-8624.2012.01848.x

Breuer, J., Elson, M., Mohseni, M. R., & Scharkow, M. (2012, July). *Are we really only measuring media effects? Problems and pitfalls associated with the implementation and analysis of the Competitive Reaction Time Task (CRTT) in research on digital games.* Paper presented at the 17th Workshop Aggression, Luxembourg.

Brown v. Entertainment Merchants Association, 564 U.S. __, 131 S.Ct. 2729 (2011).

Bushman, B. J., & Huesmann, L. R. (2013). Twenty-five years of research on violence in digital games and aggression revisited: A reply to Elson & Ferguson (2013). *European Psychologist*. Advance online publication. doi: 10.1027/1016-9040/a000164

Bushman, B. J., Rothstein, H. R., & Anderson, C. A. (2010). Much ado about something: Violent video game effects and a school of red herring: Reply to Ferguson and Kilburn (2010). *Psychological Bulletin, 136*, 182–187. doi: 10.1037/a0018718

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49*, 997–1003. doi: 10.1037/0003-066X.49.12.997

Common Sense Media. (2013). *Media and violence: An analysis of current research*. San Francisco, CA: Author. Retrieved from http://www.commonsensemedia.org/

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2011). The general aggression model: Theoretical extensions to violence. *Psychology of Violence, 1*, 245–258. doi: 10.1037/a0023842

Donnerstein, E., Strasburger, V. C., Viner, R., & Gentile, D. A. (2013, June). *Why don't they believe us? Why are media effects on children & adolescents routinely ignored or downplayed.* Panel presented at the 63rd Annual Conference of the International Communication Association, London, UK.

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PloS ONE, 7*, e29081. doi: 10.1371/journal.pone.0029081

Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (in press). Comparing apples and oranges? Evidence for pace of action as a confound in research on digital games and aggression. *Psychology of Popular Media Culture*. doi: 10.1037/ppm0000010

Ferguson, C. J. (2009). Is psychological research really as good as medical research? Effect size comparisons between psychology and medicine. *Review of General Psychology, 13*, 130–136. doi: 10.1037/a0015103

Ferguson, C. J., Ivory, J. D., & Beaver, K. M. (2013). Genetic, maternal, school, intelligence, and media use predictors of adult criminality: A longitudinal test of the Catalyst Model in adolescence through early adulthood. *Journal of Aggression, Maltreatment & Trauma, 22*, 447–460. doi: 10.1080/10926771.2013.785457

Ferguson, C. J., & Kilburn, J. (2010). Much ado about nothing: The misestimation and overinterpretation of violent video game effects in Eastern and Western nations: Comment on Anderson et al. (2010). *Psychological Bulletin, 136*, 174–178. doi: 10.1037/a0018566

Ferguson, C. J., Muñoz, M. E., & Medrano, M. R. (2012). Advertising influences on young children's food choices and parental influence. *The Journal of Pediatrics, 160*, 452–455. doi: 10.1016/j.jpeds.2011.08.023

Ferguson, C. J., & Rueda, S. M. (2009). Examining the validity of the modified Taylor competitive reaction time test of aggression. *Journal of Experimental Criminology, 5*, 121–137. doi: 10.1007/s11292-009-9069-5

Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S., & Smith, S. M. (2008). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation? *Criminal Justice and Behavior, 35*, 311–332. doi: 10.1177/0093854807311719

Ferguson, C. J., Smith, S. M., Miller-Stratton, H., Fritz, S., & Heinrich, E. (2008). Aggression in the laboratory: Problems with the validity of the modified Taylor competitive reaction time test as a measure of aggression in media violence studies. *Journal of Aggression, Maltreatment & Trauma, 17*, 118–132. doi: 10.1080/10926770802250678

Gentile, D. A. (2013, June). *The myth of the fairness doctrine: Why both sides of the issue don't need to be represented!* Paper presented at the 63rd Annual Conference of the International Communication Association, London, UK.

Giancola, P. R., & Parrott, D. J. (2008). Further evidence for the validity of the Taylor aggression paradigm. *Aggressive Behavior, 34*, 214–229. doi: 10.1002/ab.20235

Hall, R. C. W., Day, T., & Hall, R. C. W. (2011a). A plea for caution: Violent video games, the Supreme Court, and the role of science. *Mayo Clinic Proceedings, 86*, 315–321. doi: 10.4065/mcp. 2010.0762

Hall, R. C. W., Day, T., & Hall, R. C. W. (2011b). A further plea for caution against medical professionals overstating video game violence effects. *Mayo Clinic Proceedings, 86*, 821–823. doi: 10.4065/mcp. 2011.0357

Henrich, C. C., & Shahar, G. (2013). Effects of exposure to rocket attacks on adolescent distress and violence: A 4-year longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*, 619–627. doi: 10.1016/j.jaac.2013.02.016

Huesmann, L. R. (2013, June). *Habitual exposure to media violence in childhood predicts to serious aggression and crime in late adolescence and adulthood: Evidence from four longitudinal studies.* Paper presented at the 63rd Annual Conference of the International Communication Association, London, UK.

Krahé, B. (2013). Restoring the spirit of fair play in the debate about violent video games: A comment on Elson & Ferguson (2013). *European Psychologist.* Advance online publication. doi: 10.1027/1016–9040/a000165

Krahé, B., Möller, I., Huesmann, L. R., Kirwil, L., Felber, J., & Berger, A. (2011). Desensitization to media violence: Links with habitual media violence exposure, aggressive cognitions,

and aggressive behavior. *Journal of Personality and Social Psychology, 100*, 630–646. doi: 10.1037/a0021711
Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7*, 109–117. doi: 10.1177/1745691611432343
Orue, I., Bushman, B. J., Calvete, E., Thomaes, S., de Castro, B. O., & Hutteman, R. (2011). Monkey see, monkey do, monkey hurt. Longitudinal effects of exposure to violence on children's aggressive behavior. *Social Psychological and Personality Science, 2*, 432–437. doi: 10.1177/1948550610396586
Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PloS ONE, 7*, e42510. doi: 10.1371/journal.pone.0042510
Pollard Sacks, D., Bushman, B. J., & Anderson, C. A. (2011). Do violent video games harm children? Comparing the scientific amicus curiae ''experts'' in Brown v. Entertainment Merchants Association. *Northwestern University Law Review Colloquy, 106*, 1–12.
Rothstein, H. R., & Bushman, B. J. (2012). Publication bias in psychological science: Comment on Ferguson and Brannick (2012). *Psychological Methods, 17*, 129–136. doi: 10.1037/a0027128
Savage, J. (2004). Does viewing violent media really cause criminal violence? A methodological review. *Aggression and Violent Behavior, 10*, 99–128. doi: 10.1016/j.avb.2003.10.001
Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., ... Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PloS ONE, 8*, e56515. doi: 10.1371/journal.pone.0056515
Sherry, J. L., Lucas, K., Greenberg, B. S., & Lachlan, K. A. (2006). Video game uses and gratifications as predictors of use and game preference. In P. Vorderer & J. Bryant (Eds.), *Playing video games. Motives, responses, and consequences* (pp. 213–224). Mahwah, NJ: Erlbaum.
Statens Medieråd. (2011). *Våldsamma datorspel och aggression. En översikt av forskningen 2000–2011.* [Violent computer games and aggression. An overview of the research 2000–2011]. Stockholm, Sweden: Author. Retrieved from http://www.statensmedierad.se
Strasburger, V. C., & Donnerstein, E. (2013). The new media of violent video games: Yet same old media problems? *Clinical Pediatrics*. Advance online publication. doi: 10.1177/0009922813500340
Subcommittee of Youth Violence. (2013). *Youth Violence: What we need to know*. Washington, DC: Advisory Committee to the Social, Behavioral and Economic Sciences Directorate, National Science Foundation.
Surette, R. (2013). Cause or catalyst: The interaction of real world and media crime models. *American Journal of Criminal Justice, 38*, 392–409. doi: 10.1007/s12103-012-9177-z
Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior, 1*, 163–177. doi: 10.1016/1359-1789(95)00014-3
Tedeschi, J. T., & Quigley, B. M. (2000). A further comment on the construct validity of laboratory aggression paradigms: A response to Giancola and Chermack. *Aggression and Violent Behavior, 5*, 127–136. doi: 10.1016/s1359-1789(98)00028-7
Van Looy, J., Quandt, T., Vogelgesang, J., Elson, M., Ivory, J. D., Mäyrä, F., & Consalvo, M. (2013, June). *Mapping the field of digital games research: Results of a large international survey.* Paper presented at the 63rd Annual Conference of the International Communication Association, London, UK.
Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*, 274–290. doi: 10.1111/j.1745-6924.2009.01125.x
Warburton, W. (2013). Apples, oranges and the burden of proof – putting media violence findings into context: A Comment on Elson & Ferguson (2013). *European Psychologist*. Advance online publication. doi: 10.1027/1016-9040/a000166
Wright, J. C., Huston, A. C., Reitz, A. L., & Piemyat, S. (1994). Young children's perceptions of television reality: Determinants and developmental differences. *Developmental Psychology, 30*, 229–239. doi: 10.1037//0012-1649.30.2.229

Published online December 10, 2013

Malte Elson

Department of Communication
University of Münster
Bispinghof 9–14
48143 Münster
Germany
Tel. +49 251 83-21263
Fax +49 251 83-21310
E-mail malte.elson@uni-muenster.de

Digital Games in Laboratory Experiments: Controlling a Complex Stimulus through Modding

Malte Elson and Thorsten Quandt

University of Münster

Author Note

Malte Elson, Department of Communication, University of Münster, Department of Communication, University of Münster, Bispinghof 9-14, 48143 Münster, Germany.

E-Mail: malte.elson@uni-muenster.de; Tel: +49-251-83-21263

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 240864 (SOFOGA).

The authors would like to thank Johannes Kaiser and Christoph Langenberg for their assistance in the work that led to this paper.

Digital Games in Laboratory Experiments: Controlling a Complex Stimulus through Modding

Abstract

This article is a methodological examination of standards and practices when using digital games as stimulus material in laboratory experiments, particularly media effects research. It is concerned with the common lack of clean experimental manipulation and proper stimulus control in games research practices. We first discuss how scholars have addressed this issue in the past and then introduce game modifications ("modding") as a viable alternative. Successful applications of modding in experiments are outlined, and followed by a brief overview of modding tools readily available for research purposes. We demonstrate that modding is a method providing researchers with the necessary tools for powerful variable manipulations and operationalizations. At the same time, researchers maintain a thorough control over their stimulus materials, and are able to create proper experimental and control groups. Moreover, it increases studies' internal validity and replicability without necessarily impairing their ecological validity.

Keywords: Digital games; Modding; Modifications; Methods; Experiments

Digital Games in Laboratory Experiments: Controlling a Complex Stimulus through Modding

Introduction

Similar to the rise in popularity of computer and video games in the general population (Entertainment Software Association, 2013) there has been a growing interest in their uses and effects in academia, particularly in social and media psychology (Washburn, 2003). Since a multitude of cognitive, emotional, and behavioral processes are involved in playing, games can be a useful tool to examine key concepts of psychology (Järvelä, Ekman, Kivikangas, & Ravaja, 2014; Washburn, 2003). Researchers have conducted many studies to investigate links between games and relevant behavioral outcomes, such as aggression, risk-taking, and prosociality (for a recent review of popular research topics involving digital games, see Ivory, 2013). Irrespective of the topic, many laboratory experiments on uses and effects of digital games share a certain pattern in their design: One group of participants plays a game featuring a particular characteristic (the independent variable), while the other group plays a different game without this characteristic (or less pronounced), during or after which dependent variables are measured. Games, being highly complex and dynamic media, introduce new challenges to the methods of media effects research, and make high demands on scholars’ capabilities (Schmierbach, 2009). A problem in the comparison of game contents in experimental conditions is that they usually differ on more dimensions than just the one of interest to the researcher.

This issue will be discussed throughout this article in two parts: In the first half, we will discuss how the complexity of digital games as a stimulus has been addressed, and the different strategies by scholars to manipulate and control them for their research purposes.

The second presents game modifications (“modding”) as a viable solution to some of the issues discussed in the first part. We present a systematic review investigating the prevalence of modding as a technique in games research. This is followed by an outline of successful applications of game modifications in psychological experiments. We provide by a brief overview of modding tools readily available. Finally, some practical advice is given for the modus operandi of creating a mod for research purposes.

Digital Games: A multivariate stimulus

The key characteristic of the experimental method (in psychology as in other sciences) is the variation of conditions (independent variables) under scrutiny, and the simultaneous control of all other variables (confounds) that might be of relevance to the measurements (dependent variables) taken. Confounds must be controlled so they do not interfere with any effect that should be explained exclusively by the manipulation. And while this assertion is being taught in any ordinary introductory psychology class, it has serious consequences for research on and with digital games.

While it bears a great convenience, and certainly a convincing face validity, to divide games into two groups according to the current variable of interest (e.g., violent and nonviolent games), it should be considered that the occurrence of this variable is unlikely to be the only difference between two games that have been selected for research purposes. Those additional differences pose potentially confounding factors that might bias results if they are not controlled for. This problem is particularly intricate as there are common cooccurrences of themes, contents, and mechanisms in certain genres (Apperley, 2006) that could lead to a systematic conflation in larger bodies of research. The implied risk is that, instead of systematic and programmatic research of content and context variables, games research ultimately turns out to be a series of case studies comparing individual titles with each other.

But what are those common characteristics? For the example of violent games, Adachi and Willoughby (2011a) argue that when investigating effects on aggressiveness, scholars should consider the *difficulty*, *pace of action*, and *competitiveness* of a game as possibly relevant variables besides *violent content*. A genre that most often features displays of violence is the first-person shooter (e.g., *Counter-Strike*; Valve, 2000). These games are usually also fast-paced, likely to be played competitively against other human players, and highly demanding in terms of perception and motor abilities – not to mention that first-person shooters are always played from the first-person perspective. By contrast puzzle games (e.g., *Tetris*), popular stimuli for “control groups”, are typically nonviolent, but also rather slow-paced, usually played alone, and require cognitive efforts, such as problem-solving abilities. So when observing differences in measurements between those groups after playing, does it mean we have learned something about the effects of a particular game characteristic on human behavior? Adachi and Willoughby (2011a) fear that many studies attributing changes in outcome variables to specific types of content might be severely confounded by other contents that were not properly controlled for, or even accidentally manipulated by using different games that varied on multiple dimensions. This could, in fact, be one reason why the research on effects of violence specifically has been rather diverse (and inconclusive) in its results (Ferguson, 2013).

An example for this kind of research design can be found in the study by K. Williams (2009). Participants played either *Mortal Kombat: Deception* (Midway, 2004) or *Dance Dance Revolution Max 2* (Konami, 2003). *Mortal Kombat* is a fighting game in which the player controls a character engaged in close combat with an opponent (either controlled by another player or, as in this study, by the computer). One match usually involves several rounds of fighting in a confined space (such as an arena). *Dance Dance Revolution* on the other hand is a rhythm game in which players typically have to mimicry dancing instructions to pop songs on special mats that work as input devices. In this study, however, participants used a regular game controller instead of dance mats, and had to mimicry button sequences presented on screen. After playing they completed a hostility scale, showing that those who played *Mortal Kombat* reported significantly stronger feelings of hostility. The author concludes that “[t]his supports past evidence that exposure to violent video games, when compared to nonviolent video games, results in aggressive affect” (p. 303). While there are certainly theoretical arguments to be made for and against the effects of game violence on hostility (which we will abstain from discussing here), the lack of control over the stimulus material makes it difficult to disambiguate the effects of one variable from another.

That is not to say the study is irrelevant, or does not contribute to our understanding of media effects. We acknowledge that this experiment was probably exploratory in nature, and should be built upon in further studies. However, the many differences between fighting and rhythm games in general, and *Mortal Kombat: Deception* and *Dance Dance Revolution Max 2* in particular, do not allow conclusions about *one variable* (violence) stated with such certainty. Granted a bit of silliness here, the results would also allow us to conclude that exposure to dancing games, when compared to nondancing games, results in reduced aggressive affect (e.g., because the participants might have experienced dancing, even if just in a game, as a particularly pleasant, nonhostile activity). Other studies with similar problems in stimulus control be found in the literature. Ballard and Wiest (1996), for example, should be applauded for using in-game options of *Mortal Kombat* (Midway, 1992) to manipulate the degree of violence between conditions. Using a billiards game for another control group, however, probably introduced a number of confounding variables into the group comparisons.

An example for a study with similar research questions for which modding was used to control the stimulus material is the one by Staude-Müller, Bliesener, and Luthman (2008), whose participants either played the first-person shooter *Unreal Tournament 2003* (Epic

Games & Digital Extremes, 2003) or its modification *Team Freeze 2003* (Ootpik, 2006) in which avatars are being frozen instead of killed. Of course, this sophisticated manipulation might not fully solve problems of stimulus control, as it could be questioned whether freezing someone should be considered truly “nonviolent” (after all, like dying, freezing is not a very pleasant experience). And while we cannot reject this criticism in *absolute* terms, we would argue that it is still a highly functional approximation of a clear *relative* difference between conditions. Any outcome variable that differs between the “kill” and the “freeze” version could be attributed to the degrees of manipulation, even when the latter version does not remove *all* violence from the former (and should there be no appreciable difference between these conditions, this would be another important finding).

Suggestions and Practices

Other scholars already have lamented the lack of stimulus control and problems with adequate stimulus selection in laboratory games research and suggested different strategies to address this issue, which will be presented briefly here. After this, we will introduce modding as an alternative, or rather complementary solution to those suggested elsewhere.

Selecting the “right” games

D. Williams (2005) advocates a phenomenological approach and encourages social scientists to play games themselves. Having first-hand experience enables scholars to make informed decisions about the suitability of available games as stimulus material for research purposes. Moreover, Williams emphasizes the importance of playing games as they are naturally used by their players to ensure the ecological validity of the study. Williams concludes that “[k]nowing how a game is used is a necessary step before undertaking any research design” (p. 459). However, in academic publications only few scholars (at least in social sciences) report their own experience with the games used as stimuli. Whether or not this has been put into practice remains therefore speculative.

Stimulus control through ratings

Carnagey and Anderson (2004) suggest that "the obvious solution for future studies is to do more pilot testing or manipulation checks" (p. 9). This procedure is arguably the most common practice to control for third variables in games research. Many scholars use variants of Anderson's (1985) Video Game Rating Sheet (or similar ad-hoc scales), which includes ratings of several basic experiential variables, such as enjoyment and frustration. And while this is indeed an obvious solution, it must be pointed out that systematic research would need to identify the dimensions relevant to outcome variables first in order to allow exertion of sufficient control. Moreover, there might also be game characteristics relevant to or interacting with independent variables without having a direct effect on dependent variables themselves.

A further issue with this practice concerns the participants’ *point of reference* when games are being rated. Put simply, it has not been clarified what exactly specific games are being compared to when they are rated with regards to content or playing experience. The current practice of choosing games based on comparisons of separate ratings works under the assumption that each title is compared to all games a participant is familiar with. Another likely scenario is that the games are being compared to other titles within the same genre, or even subgenre. Not only because genres are more salient categories than the entirety of digital games, but also because cross-genre comparisons of game characteristics could prove to be difficult for game enthusiasts. For example, the sources of difficulty in a puzzle game are inherently different than those in a first-person shooter. When after 10 minutes of playing *Tetris* participants are asked to rate its difficulty, would they compare it to their experiences with sport simulations, first-person shooters, or strategy games? It must be pointed out that, as this issue has not been subjected to any empirical research, the discussion has to remain speculative at this point. Nevertheless, our assumption would be that participants would compare *Tetris* either to other puzzle games, or even only to other variants of *Tetris* they have encountered before. Other categories besides genre that might be drawn upon could be platform, mode, or milieu (for a discussion of those game categorizations see Apperley, 2006). What sounds like (and might as well be) a trivial academic problem can potentially lead to problems in stimulus control, as a numerical equality of two games in rated variables, for example *difficulty*, does not necessarily reflect actual similarity. For some studies, one rationale might be using the games as two opposed poles in one semantic differential scale (see figure 1). That way, the ratings would yield one value per relevant control variable for both games. These ratings could be used as data weights or covariates in further statistical analyses. Naturally, raters would need to play both games consecutively. This approach has its own limitations, of course, as it might not be feasible for studies in which more than two games are used. Also, it does not solve the problem of controlling for differences between games universally, but for specific types of studies or stimulus materials, it might pose an improvement.

Same game, different modes

McMahan, Ragan, Leal, Beaton, and Bowman (2011) take a different approach: They suggest employing only one game, and using its own modes or customization options for experimental manipulations and conditions. They recommend using novel, commercial off-the-shelf titles to simultaneously achieve a sufficient ecological validity in addition to high levels of control (a rare feat). There are some studies that actually used different versions of the same game. Participants in a study by Schmierbach (2010) were assigned to three different modes (solo, competitive, cooperative) of *Halo: Combat Evolved* (Bungie, 2001) instead of using three different games. This allowed Schmierbach to show that playing the game cooperatively results in reduced aggressive cognitions compared to playing solo or competitively, even with considerable amounts of violence present in all three conditions.

McMahan et al. (2011) and Järvelä et al. (2014) note, however, that the variety of customizations or modes included in many commercial titles might be restricted, or at least too limited to suit academic research questions. And, of course, picking the “right” game for a study requires obtaining a vast familiarity with the current games market, which often entails behaviors incompatible with usual working practices at academic institutions (certainly a particular reputation at the researchers’ department).

In the following, we will try to propose the method of modding as a viable approach for game researchers to effectively manipulate game variables, control other relevant characteristics, and maintain sufficient ecological validity. We want to emphasize that this approach does not necessarily rival the ones discussed above, but complements them. Scholars are still advised to play games themselves, as argued by D. Williams (2005), and to use rating sheets for manipulation checks or variables that could not be modified. First, we will present a conceptualization of modding and other relevant terms. We will then outline a brief history of modding both in entertainment and in past research, followed by a discussion of modding tools readily available to academics.

What is a “mod”?

Modding or *mod* originates from the word *modification*. According to the rather broad definition by Scacchi (2010), the term mod covers “customizations, tailorings, and remixes of game embodiments, whether in the form of game content, software, or hardware”. He identified five different types of mods (see Table 1): *(1) User interface customizations*, which include modifications of avatar appearance, the interface color palette and style, as well as functional and nonfunctional add-ons to the heads-up display; *(2) Game conversions*, either partial (smaller additions to an existing game), or total (entirely new games in terms of narrative, setting, mechanisms, or even genre); *(3) Machinima and art mods*, usually facilitating a cinematic storytelling experience, or posing some sort of visual static, dynamic, or performance art, and sometimes as a tool for exhibition purposes; *(4) Custom playing PCs*, which feature hardware modifications to maximize the computer's performance, or give it a distinctive aesthetic appeal; and *(5) Game console hacking*, which involves unlocking of game consoles' functions, applications, or services through hardware modifications (often in a legal gray area, to say the least). While technically all these types could be relevant to research involving digital games, this article focuses mostly the subtype of *partial conversions*, as those are particularly relevant to media effects researchers.

The development of mods is commonly undertaken by individuals or a small group of players, and only rarely by established game developing companies. Game publishers sometimes get involved in distributing particularly successful mods among the gaming community. However, even the most sophisticated mods require the user to have the original game in order to run them. Smaller mods consist of cosmetic additions, such as new textures for existing objects and characters (*skins*), or replace existing minor game content such as sounds or background music (*ambience*). More advanced mods sometimes introduce new three-dimensional items (*meshes*; e.g., weapons, armor) that can heavily affect gameplay, or they allow a game to be played in entirely new levels or areas (*maps*). Extensive mods add new quests and chapters to a game's story (including artistic assets, narrative elements, characters, dialogues), and additional game modes (particularly multi-player variants). Generally, mods can simply be installed by pasting the files in a specific folder of the respective game.

It is not uncommon for PC games to be designed with modifications in mind. Many game publishers not only allow players to alter and publish content, but also provide them with the necessary modding tools, such as source code, level editors, compilers, and rich documentation to assist mod makers and ensure a generally high quality of published mods.

With these resources facilitating high quality, it is even possible for a game mod to become equally or more popular than the original game it derived from.

A brief history of modding

While there are different opinions on what could be regarded as the first mod, Au (2002) considers *Castle Smurfenstein* (Johnson & Nevins, 1983), a parody of the classic *Castle Wolfenstein* (Muse Software, 1981) replacing Nazis with Smurfs, to be one of the earliest cornerstones in modding history. However, it was arguably the release of the first-person shooter *Doom* (id Software, 1993) and the editing tools that came with it that gave rise to today's modding scene. Many other developers, but particularly id Software heavily promoted and supported modding as it was essentially additional free content that attracted more players to purchase their games. Many other games, such as *Quake II* (id Software, 1997), *Unreal* (Epic MegaGames & Digital Extremes, 1998), and *Half-Life* (Valve, 1998), followed this example and included powerful editing tools in the game software all of which sparked an enormous amount of mods (small and big) that were at first distributed through complimentary CDs in game magazines. In 2000, Valve acquired the rights to the fan-made mod *Counter-Strike* and published it as a stand-alone title. This mod turned *Half-Life,* a single-player story-driven first-person shooter set in a science-fiction dystopia, into a tactical multi-player game in which terrorists combat Special Forces round after round. *Counter-Strike* remains perhaps the most successful mod, until today – and propelled modders to the big stages of the gaming industry, making it a potentially profitable activity (Postigo, 2007). Players of *The Sims* (Maxis, 2000), and its sequels, were suddenly able to download countless new household items, wallpapers, and dresses. Some of those modders turned their back on the traditional free-for-all spirit of the modding community, and – with considerable success – gave price tags to their creations. One recent noteworthy example is the mod *DayZ* (Hall, 2012) that turned the military simulation *ARMA II* (Bohemia Interactive, 2009) into a multiplayer open-world zombie survival game. Its release put the 3 year older original *ARMA II* back on top of the game sales charts. Numerous successful and noteworthy mods could not be mentioned here due to space limitations. For a more detailed history of mods see Champion (2012).

Modding in digital games research

To investigate the prevalence and purposes of modding in research, we conducted a brief two-step systematic literature search. In a third step we present some of the findings obtained through the literature search in a narrative review of modding in experimental work.

Step 1: Database search. We used EBSCOHost to search the databases *Academic Search Premier, PsycARTICLES, PsycINFO,* and *Communication & Mass Media Complete* for all peer-reviewed entries through September 2013 using the following three terms in *All Text (TX)*: *modding*; *video game** and *mods; video game** and *modif**. From this search we retrieved a total of 52 publications that either dealt with the topic of modding specifically or utilized modding techniques for research purposes. We coded the methodological approaches and themes or topics of these publications to get a rough estimate of the most prevalent contents (see Table 2 and Appendix A). Entries could receive multiple codes when their scope was broader, or they covered several topics in particular.

35% of all articles presented empirical data, while only 10% had a proper experimental design (with effects of modified independent variables on dependent variables). 14% had a detailed guide for how to mod contents for certain purposes (mostly for professional practice, e.g. teachers). Topic-wise, the majority (42%) of the entries dealt with modding as a means of cultural production. 10% considered modding as a form of art, or a technique to create art. At least 23% presented either theoretical or practical ideas for modding in an educational context, for example having students create their own mod (teaching programming languages), or use existing mods to engage them in learning about its contents (e.g., history). A third of the papers dealt with labor practices of modding, its economic value and implications, or its relations to the commercial gaming industry. 19% focused on legal aspects of modding in their creation and distribution, focusing particularly on issues of copyright. Finally, 15% scrutinized personological and motivational variables of members in the modding community.

This left us with a result that was rather unsatisfying. The larger part of the studies on modding we obtained was not even empirical, much less experimental in nature. Of course, one possible (and likely) reason for this finding could be that scholars make use of modding techniques without referring to them as such or using the terminology more common among game designers than social scientists. Therefore, we expanded our literature review in a second step.

Step 2: Journal search. We reviewed all articles (*N* = 4,160) published in the last ten volumes of six of the most relevant journals in the field of empirical media effects research: *Communication Research* (Sage), *Computers in Human Behavior* (Elsevier), *Cyberpsychology, Behavior, and Social Networking* (Mary Ann Liebert), *Human Communication Research* (Wiley), *Journal of Communication* (Wiley), and *Media Psychology* (Taylor & Francis). These journals were selected because they regularly publish empirical (particularly experimental) work on digital game effects, and they are all ranked in the Thomson Reuters Social Sciences Citation Index (SSCI). To keep the focus on research to which modding could be applied as a means of stimulus creation, manipulation, or control, we identified a subsample of *n* = 145 articles in which at least one game was used as stimulus material for an experiment (see Table 3 and Appendix B).

As suspected, there were several studies in which researchers modified games or created new maps and objects for existing games without referring to this technique as "modding" (or naming modding tools that were used). Of the 145 studies that employed digital games as stimuli, 26 (18%) could be identified in which materials were modded by manipulating contents so that the games would be more suited for their research questions (e.g., varying contents to create conditions, or removing unwanted contents to exert greater stimulus control). Naturally, not all of the remaining 119 studies would have substantially benefited from using modding techniques *per se*, as the research questions were varied greatly. Some studies were, for example interested in presentation modes (screen vs. head-mounted display), or differential effects of input devices (controller vs. natural mapping, e.g. steering wheel). We therefore would like to remain rather conservative with specific suggestions where modding might have been advisable. In 42 studies (29%) at least one independent variable was manipulated by using two or more completely different commercial titles (potentially diminishing internal validity), and for 28 studies (19%) entirely new games were created as stimulus materials (potentially diminishing external validity). At least in these cases, modding one game instead, or using different modes of the same game (McMahan et al., 2011) might have been viable alternatives. Naturally, this review should not be taken as representative of the whole field of digital games effects research, as there is a large number of other journals that also publishes experimental games research (such as, for example, *Psychology of Popular Media Culture*). Given the prestigious ranking in the SSCI that all six selected top-tier journals had, however, we still consider these findings to be important.

From these two steps we learned that only a minority of scholars use modding for their experiments, although they unsurprisingly describe the process with the vocabulary of their home discipline, and not with the terminology coined by the modding community. The majority of publications that was found to explicitly use the term modding in step 1 either stems from fields with less empirically-driven research traditions (e.g., the humanities) and conceived modding as part of modern culture, or took a more applied perspective on modding, e.g. as an instrument for educators and other practitioners.

Step 3: Narrative review. As this manuscript is mostly concerned with modding as a tool for experimental research, we now briefly present a selected sample of studies that successfully utilized this method in laboratory studies. With due care it can be asserted that one of the earliest examples can be found in Case, Ploog, and Fantino (1990). The authors created a total of six variants of the 1971 text-based strategy game *Star Trek* (Mayfield, 1971) by simplifying or restricting existing in-game commands and adding new ones. Thus, Case et al. tested two competing hypotheses (conditioned reinforcement vs. uncertainty reduction), and tried to convince other researchers of the benefits of user computer games in experimental psychology. More than 15 years later, Frey, Hartig, Ketzel, Zinkernagel, and Moosbrugger (2007) advocated game modding as a tool to study general human behaviors in a similar fashion by creating a mod for *Quake III Arena* (id Software, 1999) to measure navigation performance. They conclude that modding is a promising and inexpensive way to administer psychological experiments (see also de Kort, IJsselsteijn, Kooijman, & Schuurmans, 2003). For example, Frey, Blunk, and Banse (2006) used a mod for the assessment of behavioral interaction patterns (inter-avatarial distance, frequency of gaze) and relationship satisfaction of romantic couples (see also Schönbrodt & Asendorpf, 2011).

The great interest of media effects researchers in aggression-related outcomes has elicited several sophisticated experimental investigations utilizing mods. To study the effects of violent digital games on cardiovascular responses and aggressive behavior, Elson, Breuer, Van Looy, Kneer, and Quandt (2013) assigned their participants to play one of four versions of the first-person shooter *Unreal Tournament 3* (Epic Games, 2007): normal-paced (default speed level) vs. fast-paced (speed level at 140%), violent (wielding a grenade launcher) vs. nonviolent (wielding a toy nerf gun). Hartmann, Toz, and Brandon (2010) created two mods of *Operation Flashpoint: Cold War Crisis* (Bohemia Interactive, 2001) to assess the effects of (un)justified violence on feelings of guilt: Their participants were either playing UN soldiers attempting to shut down a torture camp, or paramilitary forces defending the camp and continuing the cruelty (see also Hartmann and Vorderer, 2010). Further examples of studies on aggression or moral disengagement utilizing mods can be found in Bluemke, Friedrich, and Zumbach, (2010), Carnagey and Anderson (2005), Chittaro and Sioni (2012); Hartmann and Vorderer (2010), and Mohseni (2013).

Other research areas use mods to manipulate experiential variables during play. An advanced mod of *Half-Life 2* (Valve, 2004a) can be found in Dekker and Champion (2007), which allowed the researchers to record and transfer biometric information online through finger sensors into the game. Their aim was to improve the user experience by adaptively modifying the game during play according to specific physiological responses (see also Champion & Dekker, 2011). Klimmt, Hartmann, and Frey (2007) created two modifications of a Breakout-style game to test whether reduced control (higher difficulty) and reduced effectance (unresponsive input device) would decrease game enjoyment.

Modding: A lab in a lab

Having presented several successful applications of the modding in experimental research, this last section provides a brief overview of the modding tools currently available. While it cannot replace the study of elaborate tutorials, it aims to generate a starting point, a rough idea of which tools might and might not suit researchers' needs. Using the example of three case studies, Laukkanen (2005) provides a slightly dated albeit helpful introductory text about general characteristics of modding tools and other resources (such as modding communities).

Generally speaking, mods can be a very elegant way to operationalize research questions, which means that, first of all, researchers must be very clear about their key variables of interest. Once clear hypotheses have been established, it can be determined whether modding can help in testing them, or whether other approaches might be more promising. As a rule of thumb, modding is effective in research designs when it comes to game characteristics and contents (such as game settings, the presence or absence of certain variables, or the looks and properties of objects and avatars), but less effective when it comes to structural aspects (game modes, basic game mechanisms). For example, it is relatively easy to change the looks and behaviors of opponents in a first-person shooter, but it is hard (and inefficient) to turn it into a 2D puzzle game. Picking the right game is key, as it might reduce the amount of necessary manipulations, which saves time and maintains external validity. The list by Mohseni, Elson, Pietschmann, and Liebold (2014) provides a comprehensive overview of modable games and their specific modding tools (and how to obtain them), and might be a good starting point for researchers.

The next step is to select the tools required for the operationalization. Similar to the amount and extent of variables researchers want to manipulate, they should be economical in their decision which modding tools they use to avoid the problem of “cracking a nut with a sledgehammer”. It is very possible that one does not actually need to use any modding software at all, as games often allow a large variety of settings in their own options menus. In other cases, someone else might already have created and published a mod that suits the specific needs, and can be easily installed like a regular game. The following section is meant to give a brief description of available modding tools and resources.

Mod DB

Creating a small mod for the first time(s) can be cumbersome and laborious. Fortunately, the modding community often releases their work to the public for free. It is very likely that researchers will find that someone has already been working on a mod that can be used for their purposes. Mod DB (moddb.com) is the largest modding-related website and community on the internet, hosting a total of 11,165 different mod projects[1], of which more

[1] Retrieved on March 8, 2014.

than half have been released and are readily available. Users can search Mod DB's database by platform, theme, genre, and specific title. Most mods hosted are high-quality large-scale projects, being either advanced partial or full conversions. Scholars who consider such modifications to the original game to be too substantial for their research needs and worry about internal validity issues might want to check Mod DB's 11,812 *add-ons*[1]. These are smaller mods such as new textures, objects, or maps, and, thus, represent mods that might be suitable to specific research questions. Although many mods are hosted on Mod DB, researchers might find even more on other modding-related databases (such as GameBanana.com) or websites dedicated to specific games.

FPS Creator

FPS Creator (The Game Creators, 2005) is a program that allows creating first-person shooters without requiring any knowledge of 3D modeling or programming. At a price similar to a new game it comes with several pre-designed themes including objects and avatars to freely design different interiors or exteriors. It can be modularly extended by purchasing further theme packs which are frequently published on the developer's website. By varying the theme, one can easily create two or more slightly different versions of the same game. Advanced users of *FPS Creator* can freely import objects and textures they created with other programs, and change many aspects of the pre-designed mechanisms (behaviors, events) with the built-in script language. However, although it has been updated regularly, the graphics and mechanisms it includes must be described as outdated. *FPS Creator*-created games may have a certain nostalgic charm, but they can hardly compete with modern games that players are used to. Therefore, researchers should consider whether they are willing to trade convenience for external validity (McMahan et al., 2011). The current version only runs on Microsoft Windows. Notably, *FPS Creator* is a program to create games played from a first-person perspective (not necessarily shooters), but the developer The Games Creators offers similar development tools for other genres (e.g., massively multi-player online role-playing games).

Garry's Mod

Similar to *FPS Creator*, *Garry's Mod* (Facepunch Studios, 2006) technically is not a modding tool, but – lacking a better term – a sandbox physics game. However, the term "game" is definitely misleading, as it has no objectives or scores, and, like in a sandbox, the "player" can use its tools for any purpose (e.g., research!). Users can call any object from any game based on the *Source* game engine, as well as third-party content into *Garry's Mod* and manipulate them with two main tools: the *Physics Gun*, which is used to change spatial properties of objects (position, rotation), and the *Tool Gun*, which combines and attaches different objects, or is used to create winches and gear-wheels. Avatars (ragdolls) can also be imported and users can manipulate their joints, poses, and expressions. *Garry's Mod* even supports unconventional input devices (such as Microsoft Kinect). In simple (or rather oversimplistic) terms, it is a drag-and-drop level editor in which the laws of physics can be bent at will without requiring any programming skills. The vast amount of different objects that can be used should make it possible to create any specific "world" that the researcher needs. *Garry's Mod* has a large community that offers their own works for free, providing some good starting points for beginners. It runs on Windows and Mac, and is available on *Steam* (Valve, 2003) for approximately $10.

Source SDK

Source SDK (Valve, 2004b) is a software development kit to create mods for *Source*-based games, and is available free of charge for *Steam* users. It contains three applications: *Hammer Editor*, a map creation tool; *Model Viewer*, to create and manipulate object and effect properties; and *Face Poser*, to generate facial expressions, gestures, and movements. Like *Garry's Mod*, *Source SDK* can utilize maps, textures, and objects of any game based on the *Source* engine, meaning that a vast array of themes and accordingly-themed objects is at the researcher's disposal. *Source SDK* comes with a base set of items, and every purchased *Source*-based game adds new ones to it. This means that not only the modding software itself is constantly updated, but that new items are added to it quite frequently. In addition, as *Source* is one of the most popular engines for game developers and modders, there is additional content and resources available on modding-dedicated websites. That way, researchers can easily add, move, or replace specific contents of an existing and fully developed commercial game (or a map thereof) without necessarily imposing on the game's quality or ecological validity. *Source SDK* is available for Windows, Max, and Linux, as well as Xbox 360 and PlayStation 3.

Creation Kit

Creation Kit (Bethesda, 2012) is the modding tool released for the open-world role-playing game *The Elder Scrolls V: Skyrim* (Bethesda, 2011). It is complimentary for *Steam* users who bought *Skyrim*. *Creation Kit* allows users to add or manipulate objects or items, graphical effects, character behavior, dialogue scenes, quests and events, and interior or exterior environments. Since one major feature of *Skyrim* is its open world (meaning that there is one large world map instead of many separate ones), heavy map editing is not only laborious, but might also limit the overall game enjoyment drastically. However, manipulating quests might be of particular interest for researchers, as they easily allow "framing" of in-game behaviors or tasks (e.g., acting for good vs. evil). While doing so, *Creation Kit* tries to stay as visual as possible, meaning that for basic manipulations programming or scripting skills are not necessarily required. Since *Skyrim* is set in a Dark Ages fantasy world, most of the available items are medieval-themed, which might be too limiting for scholars, depending on their research interest. However, with a growing modding community, more and more downloadable items are of a different nature. Researchers might find the video tutorials officially released by the game developer helpful to get started with *Creation Kit*. Currently it only runs on Windows, although users have reportedly made it work on Mac and Linux.

Conclusions

This paper discussed the mismatch between requirements of laboratory experiments to be standardized and controlled, and the use of digital games, being interactive and complex media, as experimental stimuli. Researchers usually make use of manipulation checks or small pre-studies to address this problem. However, as has been stated previously (D. Williams, 2005), this method has its own limitations. After all, a manipulation check can only be as good as the experimental manipulation and control. We suggested a slight improvement for self-report controls, and then introduced game modding as an alternative method. Ideally, we would recommend taking all three steps that we presented in this manuscript: Getting first-hand experience to select the “right” game, applying modifications as necessary for the research questions, and then evaluating these new variants with self-report measures. Reporting basic information (the game selected, the parts modified, and the tools used) in publications not only increases their replicability, but can also guide other less experienced scholars who might encounter problems that have already been solved.

Our final remarks are fairly straightforward: If researchers should decide that they might want to try out modding for their purposes, they need to “get their hands dirty”. While modding tools certainly look a bit overwhelming, the best way to learn how to separate advanced from simple tools (which are likely to be completely sufficient for researchers) is to install modding software, open a sample map, and simply “play around”. Try to place a tree, add an ammunitions depot, or replace a door with a solid wall. The entry threshold into modding might seem comparably high, but we hope to have demonstrated that it is worthwhile: It provides researchers with the necessary tools for powerful manipulations. At the same time, researchers maintain a thorough control over their stimulus materials, and are able to create proper experimental and control groups. Moreover, it increases studies' internal validity and replicability without necessarily impairing their ecological validity. In need of advice, there are extensive wikis and tutorials for all modding tools available, and the modding community is eager to help those who are new to their field. Virtually all modding tools have comprehensible graphical user interfaces and do not require programming skills. Finally, the successful applications of modding by other researchers presented in this paper, as well as available mods originally created for recreational purposes, can be a good starting point for further research.

References

Adachi, P. J. C., & Willoughby, T. (2011). The effect of violent video games on aggression: Is it more than just the violence? *Aggression and Violent Behavior*, *16*(1), 55–62. doi:10.1016/j.avb.2010.12.002

Anderson, C. A. (1985). *Video game rating sheet*. Retrieved from http://www.psychology.iastate.edu/faculty/caa/Scales/VGRating.pdf

Apperley, T. H. (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation & Gaming*, *37*(1), 6–23. doi:10.1177/1046878105282278

Au, W. J. (2002, April 16). Triumph of the mod. *Salon*. Retrieved from http://www.salon.com

Ballard, M. E., & Wiest, J. R. (1996). Mortal Kombat (tm): The effects of violent videogame play on males' hostility and cardiovascular responding. *Journal of Applied Social Psychology*, *26*(8), 717–730. doi:10.1111/j.1559-1816.1996.tb02740.x

Bethesda. (2011). The Elder Scrolls V: Skyrim. Rockville, MD: Bethesda Softworks.

Bethesda. (2012). Creation Kit. Retrieved from http://www.creationkit.com

Bluemke, M., Friedrich, M., & Zumbach, J. (2010). The influence of violent and nonviolent computer games on implicit measures of aggressiveness. *Aggressive Behavior*, *36*(1), 1–13. doi:10.1002/ab.20329

Bohemia Interactive. (2001). Operation Flashpoint: Cold War Crisis. Southam, UK: Codemasters.

Bohemia Interactive. (2009). ARMA II. Milan, Italy: 505 Games.

Bungie. (2001). Halo: Combat Evolved. Redmond, WA: Microsoft Game Studios.

Carnagey, N. L., & Anderson, C. A. (2004). Violent video game exposure and aggression: A literature review. *Minerva Psichiatrica*, *45*(1), 1–18.

Carnagey, N. L., & Anderson, C. A. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition, and behavior. *Psychological Science*, *16*(11), 882–889. doi:10.1111/j.1467-9280.2005.01632.x

Case, D. A., Ploog, B. O., & Fantino, E. (1990). Observing behavior in a computer game. *Journal of the Experimental Analysis of Behavior*, *54*(3), 185–199. doi:10.1901/jeab.1990.54-185

Champion, E. (Ed.). (2012). *Game mods: Design, theory and criticism*. Pittsburgh, PA: ETC Press.

Champion, E., & Dekker, A. (2011). Biofeedback and virtual environments. *International Journal of Architectural Computing*, *9*(4), 377–396. doi:10.1260/1478-0771.9.4.377

Chittaro, L., & Sioni, R. (2012). Killing non-human animals in video games: A study on user experience and desensitization to violence aspects. *PsychNology*, *10*(3), 215–243. Retrieved from http://www.psychnology.org

De Kort, Y. A. W., IJsselsteijn, W. A., Kooijman, J., & Schuurmans, Y. (2003). Virtual laboratories: Comparability of real and virtual environments for environmental psychology. *Presence: Teleoperators and Virtual Environments*, *12*(4), 360–373. doi:10.1162/105474603322391604

Dekker, A., & Champion, E. (2007). Please biofeed the zombies: Enhancing the gameplay and display of a horror game using biofeedback. In *3rd Conference of the Digital Games Research Association*. Tokyo, Japan. Retrieved from http://www.digra.org

Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (2013). Comparing apples and oranges? Evidence for pace of action as a confound in research on digital games and aggression. *Psychology of Popular Media Culture*. Advance online publication. doi:10.1037/ppm0000010

Entertainment Software Association. (2013). 2013 Sales, demographic and usage data. Essential facts about the computer and video game industry. Retrieved from http://www.theesa.com

Epic Games. (2007). Unreal Tournament 3. Chicago, IL: Midway Games.

Epic Games, & Digital Extremes. (2003). Unreal Tournament 2003. New York, NY: Atari.

Epic MegaGames, & Digital Extremes. (1998). Unreal. New York, NY: GT Interactive.

Facepunch Studios. (2006). Garry's Mod 10. Retrieved from http://www.garrysmod.com

Ferguson, C. J. (2013). Violent video games and the Supreme Court: Lessons for the scientific community in the wake of Brown v. Entertainment Merchants Association. *American Psychologist*, *68*(2), 57–74. doi:10.1037/a0030597

Frey, A., Blunk, H. A., & Banse, R. (2006). Psi-Land: Paarinteraktionsforschung in einer virtuellen Umgebung. [Couple interaction research in a virtual environment]. *Zeitschrift Für Sozialpsychologie*, *37*(3), 151–159. doi:10.1024/0044-3514.37.3.151

Frey, A., Hartig, J., Ketzel, A., Zinkernagel, A., & Moosbrugger, H. (2007). The use of virtual environments based on a modification of the computer game Quake III Arena® in psychological experimenting. *Computers in Human Behavior*, *23*(4), 2026–2039. doi:10.1016/j.chb.2006.02.010

Hall, D. (2012). DayZ. Prague, Czech Republic: Bohemia Interactive.

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, *13*(4), 339–363. doi:10.1080/15213269.2010.524912

Hartmann, T., & Vorderer, P. (2010). It's okay to shoot a character: Moral disengagement in violent video games. *Journal of Communication*, *60*(1), 94–119. doi:10.1111/j.1460-2466.2009.01459.x

id Software. (1993). Doom. Mesquite, TX: id Software.

id Software. (1997). Quake II. Santa Monica, CA: Activision.

id Software. (1999). Quake III Arena. Santa Monica, CA: Activision.

Ivory, J. D. (2013). Video games as a multifaceted medium: A review of quantitative social science research on video games and a typology of video game research approaches. *Review of Communication Research*, *1*(1), 31–68. Retrieved from http://www.rcommunicationr.org

Järvelä, S., Ekman, I., Kivikangas, J. M., & Ravaja, N. (2014). A practical guide to using digital games as an experiment stimulus. *Transactions of the Digital Games Research Association*, *1*(2), 85–116.

Johnson, A., & Nevins, P. (1983). Castle Smurfenstein. Retrieved from http://www.evl.uic.edu/aej/smurf.html

Klimmt, C., Hartmann, T., & Frey, A. (2007). Effectance and control as determinants of video game enjoyment. *CyberPsychology & Behavior*, *10*(6), 845–847. doi:10.1089/cpb.2007.9942

Konami. (2003). Dance Dance Revolution Max 2. Tokyo, Japan: Konami.

Laukkanen, T. (2005). *Modding scenes: Introduction to user-created content in computer gaming*. Tampere, Finland: University of Tampere, Hypermedia Laboratory. Retrieved from http://tampub.uta.fi/bitstream/handle/10024/65431/951-44-6448-6.pdf

Maxis. (2000). The Sims. Redwood City, CA: Electronic Arts.

Mayfield, M. (1971). Star Trek. Retrieved from http://www.dunnington.u-net.com/public/startrek/

McMahan, R. P., Ragan, E. D., Leal, A., Beaton, R. J., & Bowman, D. A. (2011). Considerations for the use of commercial video games in controlled experiments. *Entertainment Computing*, *2*(1), 3–9. doi:10.1016/j.entcom.2011.03.002

Midway. (1992). Mortal Kombat. Chicago, IL: Midway.

Midway. (2004). Mortal Kombat: Deception. Chicago, IL: Midway.

Mohseni, M. R. (2013). *Virtuelle Nothilfe. Ein Experiment zum Effekt von virtueller Hilfe, Gewalt und Nothilfe auf Hilfe- und Gewaltverhalten. [An experiment on the effect of virtual helping, violence, and emergency assistance on helping and violent behavior.]* (Doctoral dissertation). University of Osnabrück, Germany.

Mohseni, M. R., Elson, M., Pietschmann, D., & Liebold, B. (2014). Modding for digital games research. Retrieved from http://www.medkom.tu-chemnitz.de/modding/

Muse Software. (1981). Castle Wolfenstein. Baltimore, MD: Muse Software.

Ootpik. (2006). Team Freeze 2003. Retrieved from http://www.moddb.com/mods/team-freeze-2003/

Postigo, H. (2007). Of mods and modders: Chasing down the value of fan-based digital game modifications. *Games and Culture*, *2*(4), 300–313. doi:10.1177/1555412007307955

Scacchi, W. (2010). Mods, modders, modding, and the mod scene. *First Monday*, *15*(5). Retrieved from http://www.firstmonday.org

Schmierbach, M. (2009). Content analysis of video games: Challenges and potential solutions. *Communication Methods and Measures*, *3*(3), 147–172. doi:10.1080/19312450802458950

Schmierbach, M. (2010). “Killing spree”: Exploring the connection between competitive game play and aggressive cognition. *Communication Research*, *37*(2), 256–274. doi:10.1177/0093650209356394

Schönbrodt, F. D., & Asendorpf, J. B. (2011). Virtual social environments as a tool for psychological assessment: Dynamics of interaction with a virtual spouse. *Psychological Assessment*, *23*(1), 7–17. doi:10.1037/a0021049

Staude-Müller, 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*(1), 41–50. doi:10.1024/1421-0185.67.1.41

The Game Creators. (2005). FPS Creator. Retrieved from http://www.thegamecreators.com

Valve. (1998). Half-Life. Oakhurst, CA: Sierra Entertainment.

Valve. (2000). Counter-Strike. Kirkland, CA: Valve, Sierra Studios.

Valve. (2003). Steam. Retrieved from http://store.steampowered.com

Valve. (2004a). Half-Life 2. Bellevue, WA: Valve, Sierra Entertainment.

Valve. (2004b). Source SDK. Retrieved from http://source.valvesoftware.com

Washburn, D. A. (2003). The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers*, *35*(2), 185–193. doi:10.3758/BF03202541

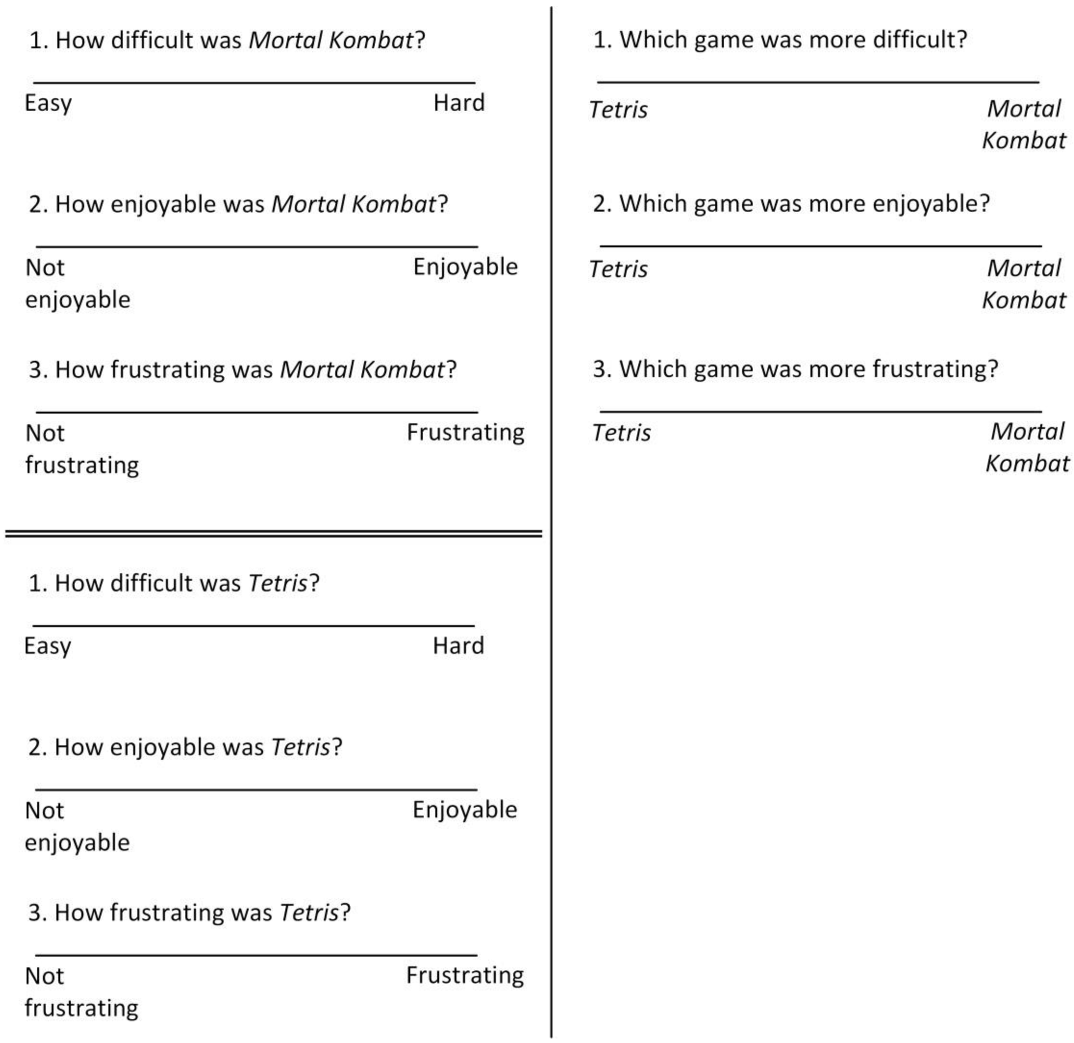
Williams, D. (2005). Bridging the methodological divide in game research. *Simulation & Gaming*, *36*(4), 447–463. doi:10.1177/1046878105282275

Williams, K. D. (2009). The effects of frustration, violence, and trait hostility after playing a video game. *Mass Communication and Society*, *12*(3), 291–310. doi:10.1080/15205430802461087

Figures

Figure 1.

Control variable ratings with the examples of *Mortal Kombat* and *Tetris*. On the left, the first three items from Anderson's (1985) Video Game Rating Sheet. On the right, the same items as semantic differentials with the two games as opposite poles.



Tables

Table 1.

The five mod types by Scacchi (2010).

Types	Classes
User interface customizations	Avatar customization, Color palette and style, interface add-ons
Game conversions	Partial and total conversions
Machinima and art mods	Cinematic storytelling, visual art, exhibition purposes
Custom playing PCs	Maximizing performance, aesthetic appearance
Game console hacking	Unlocking of functions, applications, services

Table 2.

Methodological and topical codes in the database search (step 1).

Code	n (%)
Empirical	18 (35%)
Experimental	5 (10%)
How to	7 (14%)
Cultural production	22 (42%)
Arts	5 (10%)
Labor and economy	17 (33%)
Education	12 (23%)
Legal	10 (19%)
Modders	8 (15%)
Other	5 (10%)

N = 52. Multiple codes possible.

For a full list of all references, see Appendix A.

Table 3.

Articles retrieved in the journal search (step 2).

Journal	# Articles	# Experiments	# Modding
CRX	342	4	0
CHB	1,813	51	9
CYBER	1,032	54	9
HCR	249	4	2
JoC	492	7	1
MP	232	25	5
Total	4,160	145	26

CRX = Communication Research; CHB = Computers in Human Behavior; CYBER = Cyberpsychology, Behavior, and Social Networking; HCR = Human Communication Research; JoC = Journal of Communication; MP = Media Psychology

Articles = Total number of articles published from 01/2003 through 09/2013

Experiments = Number of experiments with games as stimulus material

Modding = Number of studies using modding techniques to manipulate independent variables or control confounding factors

For a full list of all references, see Appendix B.

Appendix

Appendix A.

Literature search results.

Publication	Em	Ex	HT	Cp	Ar	La	Ed	Le	Mo	Ot
Anonymous (2012)								X		
Apperley and Jayemane (2012)				X						
Bratich and Brush (2011)				X						
Cannon (2006)					X					
Carlson and Corliss (2007)										X
Charrieras and Roy-Valex (2008)	X			X		X			X	
Clyde and Thomas (2008)			X				X			
Clyde and Wilkinson (2010)							X			
Coleman and Dyer-Witheford (2007)				X		X		X		
Corbett and Wade (2005)			X				X			
de Kosnik (2009)				X		X				
de Zwart and Lindsay (2010)								X		
Deuze (2006)				X						
Fanning (2006)			X				X			
Gershenfeld (2011)							X			
Grimes and Feenberg (2009)				X						
Hartmann, Toz, and Brandon (2010)	X	X								X
Hartmann and Vorderer (2010)	X	X								X
Harwood (2011)				X	X					
Hedberg and Brudvik (2008)							X			
Herman, Coombe, and Kaye (2006)						X		X		
Hjorth (2010)				X		X				
Hong and Chen (2013)	X					X				
Jansz and Theodorsen (2009)	X								X	
Jayemane (2009)					X					
Jöckel, Will, and Schwarzer (2008)						X				
Johnson (2009)						X		X		
Jones (2005)				X	X					
Kawashima (2010)				X				X		
Kerr (2007)				X				X		
McMichael (2007)	X		X				X			
Moore (2009)				X		X				
Moshirnia (2007)	X	X	X				X			
Nieborg and van der Graaf (2008)				X		X		X		
Owens (2010)	X					X			X	
Poor (2013)	X					X			X	
Postigo (2003)				X		X			X	
Postigo (2007)	X					X			X	
Postigo (2008)	X			X		X		X		
Riha (2012)			X	X	X					
Shah (2007)								X		

Short (2011)			X				X			
Simons (2007)				X		X				
Sinnreich and Gluck (2008)	X			X						
Šisler (2009)				X						
Sotamaa (2010)	X					X			X	
Squire, DeVane, and Durga (2008)	X						X			
Staude-Müller, Bliesener, and Luthman (2008)	X	X								X
Stowell and Shelton (2008)	X						X			
Underberg (2008)				X			X			
Wirman (2012)	X			X					X	
Young, Sutherland, and Cole (2011)	X	X								X

Em = Empirical; Ex = Experimental; Ht = How to; Cp = Cultural production; Ar = Arts; La = Labor and economy; Ed = Education; Le = Legal; Mo = Modders; Ot = Other

References

Anonymous. (2012). Spare the mod: In support of total-conversion modified video games. *Harvard Law Review*, *125*(3), 789–810.

Apperley, T. H., & Jayemane, D. (2012). Game studies' material turn. *Westminster Papers in Communication and Culture*, *9*(1), 5–25.

Bratich, J. Z., & Brush, H. M. (2011). Fabricating activism: Craft-work, popular culture, gender. *Utopian Studies*, *22*(2), 233–260. doi:10.1353/utp.2011.0043

Cannon, R. (2006). Meltdown. *Journal of Media Practice*, *7*(1), 7–23. doi:10.1386/jmpr.7.1.7/1

Carlson, R., & Corliss, J. (2007). Rubble jumping: From Paul Virilio's techno-dromology to video games and distributed agency. *Culture, Theory and Critique*, *48*(2), 161–174. doi:10.1080/14735780701723207

Charrieras, D., & Roy-Valex, M. (2008). Video game culture as popular culture? The productive leisure of video game workers of Montreal. In *58th Conference of the International Communication Association*. Montreal, Canada.

Clyde, J., & Thomas, C. (2008). Building an information literacy first-person shooter. *Reference Services Review*, *36*(4), 366–380. doi:10.1108/00907320810920342

Clyde, J., & Wilkinson, G. R. (2010). More than a game... Teaching in the gamic mode: Disciplinary knowledge, digital literacy, and collaboration. *The History Teacher*, *46*(1), 45–66.

Coleman, S., & Dyer-Witheford, N. (2007). Playing on the digital commons: Collectivities, capital and contestation in videogame culture. *Media, Culture & Society*, *29*(6), 934–953. doi:10.1177/0163443707081700

Corbett, J., & Wade, K. (2005). Player perspective: Using computer game engines for 3D cartography. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *40*(3), 113–120. doi:10.3138/8443-505Q-M8T1-1774

De Kosnik, A. (2009). Should fan fiction be free? *Cinema Journal*, *48*(4), 118–124. doi:10.1353/cj.0.0144

De Zwart, M., & Lindsay, D. (2010). Governance and the global metaverse. In D. Riha & A. Maj (Eds.), *Emerging Practices in Cyberculture and Social Networking* (pp. 65–82). Amsterdam, the Netherlands: Rodopi.

Deuze, M. (2006). Participation, remediation, bricolage: Considering principal components of a digital culture. *The Information Society*, *22*(2), 63–75. doi:10.1080/01972240600567170

Fanning, E. (2006). Game mods: Customizable learning in a K16 setting. (R. E. Ferdig, Ed.)*International Journal of Information and Communication Technology Education*, *2*(4), 15–23. doi:10.4018/jicte.2006100102

Gershenfeld, A. (2011). Leveling up from player to designer. Engaging and empowering youth through making video games. *Knowledge Quest*, *40*(1), 54–59.

Grimes, S. M., & Feenberg, A. (2009). Rationalizing play: A critical theory of digital gaming. *The Information Society*, *25*(2), 105–118. doi:10.1080/01972240802701643

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, *13*(4), 339–363. doi:10.1080/15213269.2010.524912

Hartmann, T., & Vorderer, P. (2010). It’s okay to shoot a character: Moral disengagement in violent video games. *Journal of Communication*, *60*(1), 94–119. doi:10.1111/j.1460-2466.2009.01459.x

Harwood, T. (2011). Towards a manifesto for machinima. *Journal of Visual Culture*, *10*(1), 6–12. doi:10.1177/1470412910391547

Hedberg, J. G., & Brudvik, O. C. (2008). Supporting dialogic literacy through mashing and modding of places and spaces. *Theory Into Practice*, *47*(2), 138–149. doi:10.1080/00405840801992363

Herman, A., Coombe, R. J., & Kaye, L. (2006). Your Second Life? Goodwill and the performativity of intellectual property in online digital gaming. *Cultural Studies*, *20*(2-3), 184–210. doi:10.1080/09502380500495684

Hjorth, L. (2010). The game of casual reflection: Gaming in the twenty-first century. *Metro Magazine*, *164*, 134–136.

Hong, R., & Chen, V. H.-H. (in press). Becoming an ideal co-creator: Web materiality and intensive laboring practices in game modding. *New Media & Society*. doi:10.1177/1461444813480095

Jansz, J., & Theodorsen, J. (2009). Modifying video games on web 2.0: Exploration of motives for publishing creative game content. In *59th Conference of the International Communication Association*. Chicago, IL.

Jayemane, D. (2009). The resident of Evil Creek. *Overland*, *197*, 49–56.

Jöckel, S., Will, A., & Schwarzer, F. (2008). Participatory media culture and digital online distribution - Reconfiguring the value chain in the computer game industry. *International Journal on Media Management*, *10*(3), 102–111. doi:10.1080/14241270802262419

Johnson, D. (2009). StarCraft fan craft: Game mods, ownership, and totally incomplete conversions. *The Velvet Light Trap, 64*(1), 50–63. doi:10.1353/vlt.0.0041

Jones, M. (2005). Machinima and the filmmaker's virtual immersion. *Metro Magazine, 145*, 134–137.

Kawashima, N. (2010). The rise of "user creativity" - Web 2.0 and a new challenge for copyright law and cultural policy. *International Journal of Cultural Policy, 16*(3), 337–353. doi:10.1080/10286630903111613

Kerr, A. (2007). Spilling hot coffee? Grand Theft Auto as contested cultural product. In *57th Conference of the International Communication Association*. San Diego, CA.

McMichael, A. (2007). PC Games and the teaching of history. *The History Teacher, 40*(2), 203–218.

Moore, C. L. (2009). Digital games distribution: The presence of the past and the future of obsolescence. *M/C Journal: A Journal of Media and Culture, 12*(3).

Moshirnia, A. (2007). The educational potential of modified video games. *Issues in Informing Science and Information Technology, 4*, 511–521.

Nieborg, D. B., & van der Graaf, S. (2008). The mod industries? The industrial logic of non-market game production. *European Journal of Cultural Studies, 11*(2), 177–195. doi:10.1177/1367549407088331

Owens, T. (2010). Modding the history of science: Values at play in modder discussions of Sid Meier's Civilization. *Simulation & Gaming, 42*(4), 481–495. doi:10.1177/1046878110366277

Poor, N. (in press). Computer game modders' motivations and sense of community: A mixed-methods approach. *New Media & Society*. doi:10.1177/1461444813504266

Postigo, H. (2003). From Pong to Planet Quake: Post-industrial transitions from leisure to work. *Information, Communication & Society, 6*(4), 593–607. doi:10.1080/1369118032000163277

Postigo, H. (2007). Of mods and modders: Chasing down the value of fan-based digital game modifications. *Games and Culture, 2*(4), 300–313. doi:10.1177/1555412007307955

Postigo, H. (2008). Video game appropriation through modifications: Attitudes concerning intellectual property among modders and fans. *Convergence, 14*(1), 59–74. doi:10.1177/1354856507084419

Riha, D. (2012). Machinima, creative software and education for creativity. In H. Breslow & A. Mousoutzanis (Eds.), *Cybercultures. Mediations of Community, Culture, Politics* (pp. 31–43). Amsterdam, the Netherlands: Rodopi.

Shah, S. R. (2007). Modding the web: Secondary liability under copyright and web modification software in a post-Grokster world. *Texas Law Review, 85*(3), 703–739.

Short, D. (2011). Teaching scientific concepts using a virtual world - Minecraft. *Teaching Science, 58*(3), 55–58.

Simons, I. (2007). Do it yourself: The videogame as creative tool. *Third Text, 21*(5), 627–632. doi:10.1080/09528820701600087

Sinnreich, A., & Gluck, M. (2008). Niche is the new mainstream: Measuring the growth and impact of configurable culture. In *57th Conference of the International Communication Association*. San Francisco, CA.

Šisler, V. (2009). Palestine in pixels: The holy land, Arab-Israeli conflict, and reality construction in video games. *Middle East Journal of Culture and Communication*, *2*(2), 275–292. doi:10.1163/187398509X12476683126509

Sotamaa, O. (2010). When the game is not enough: Motivations and practices among computer game modding culture. *Games and Culture*, *5*(3), 239–255. doi:10.1177/1555412009359765

Squire, K. D., DeVane, B., & Durga, S. (2008). Designing centers of expertise for academic learning through video games. *Theory Into Practice*, *47*(3), 240–251. doi:10.1080/00405840802153973

Staude-Müller, 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*(1), 41–50. doi:10.1024/1421-0185.67.1.41

Stowell, T., & Shelton, B. E. (2008). Challenges, frustrations and triumphs of remixing an open source game engine for educational purposes. *TechTrends*, *52*(5), 32–37. doi:10.1007/s11528-008-0194-1

Underberg, N. M. (2008). The Turkey Maiden educational computer game. *Folklore*, *119*(2), 201–217. doi:10.1080/00155870802057009

Wirman, H. (2012). Email interviews in player research: The case of The Sims 2 skinners. *Westminster Papers in Communication and Culture*, *9*(1), 152–170.

Young, M. E., Sutherland, S. C., & Cole, J. J. (2011). Individual differences in causal judgment under time pressure: Sex and prior video game experience as predictors. *International Journal of Comparative Psychology*, *24*(1), 76–98.

Appendix B.

Literature search results. All papers marked with a * used modding for their stimulus materials.

Admiraal, W., Huizenga, J., Akkerman, S., & ten Dam, G. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, *27*(3), 1185–1194. doi:10.1016/j.chb.2010.12.013

Bailey, R., Wise, K., & Bolls, P. (2009). How avatar customizability affects children’s arousal and subjective presence during junk food-sponsored online video games. *CyberPsychology & Behavior*, *12*(3), 277–283. doi:10.1089/cpb.2008.0292

Ballard, M. E., Hamby, R. H., Panee, C. D., & Nivens, E. E. (2006). Repeated exposure to video game play results in decreased blood pressure responding. *Media Psychology*, *8*(4), 323–341. doi:10.1207/s1532785xmep0804_1

Barlett, C. P., Rodeheffer, C. D., Baldassaro, R., Hinkin, M. P., & Harris, R. J. (2008). The effect of advances in video game technology and content on aggressive cognitions, hostility, and heart rate. *Media Psychology*, *11*(4), 540–565. doi:10.1080/15213260802492018

Barlett, C. P., Vowels, C. L., Shanteau, J., Crow, J., & Miller, T. (2009). The effect of violent and non-violent computer games on cognitive performance. *Computers in Human Behavior*, *25*(1), 96–102. doi:10.1016/j.chb.2008.07.008

Belchior, P., Marsiske, M., Sisco, S. M., Yam, A., Bavelier, D., Ball, K., & Mann, W. C. (2013). Video game training to improve selective visual attention in older adults. *Computers in Human Behavior*, *29*(4), 1318–1324. doi:10.1016/j.chb.2013.01.034

Besmann, A., & Rios, K. (2012). Pals in power armor: Attribution of human-like emotions to video game characters in an ingroup/outgroup situation. *Cyberpsychology, Behavior and Social Networking*, *15*(8), 441–443. doi:10.1089/cyber.2012.0111

Blumberg, F. C., Rosenthal, S. F., & Randall, J. D. (2008). Impasse-driven learning in the context of video games. *Computers in Human Behavior*, *24*(4), 1530–1541. doi:10.1016/j.chb.2007.05.010

Botella, C., Breton-López, J., Quero, S., Baños, R. M., García-Palacios, A., Zaragoza, I., & Alcaniz, M. (2011). Treating cockroach phobia using a serious game on a mobile phone and augmented reality exposure: A single case study. *Computers in Human Behavior*, *27*(1), 217–227. doi:10.1016/j.chb.2010.07.043

Bowman, N. D., Weber, R., Tamborini, R., & Sherry, J. (2013). Facilitating game play: How others affect performance at and enjoyment of video games. *Media Psychology*, *16*(1), 39–64. doi:10.1080/15213269.2012.742360

* Brusso, R. C., & Orvis, K. A. (2013). The impeding role of initial unrealistic goal-setting on videogame-based training performance: Identifying underpinning processes and a solution. *Computers in Human Behavior*, *29*(4), 1686–1694. doi:10.1016/j.chb.2013.01.006

Burgess, M. C. R., Dill, K. E., Stermer, S. P., Burgess, S. R., & Brown, B. P. (2011). Playing with prejudice: The prevalence and consequences of racial stereotypes in video games. *Media Psychology*, *14*(3), 289–311. doi:10.1080/15213269.2011.596467

Charles, E. P., Baker, C. M., Hartman, K., Easton, B. P., & Kreuzberger, C. (2013). Motion capture controls negate the violent video-game effect. *Computers in Human Behavior*, *29*(6), 2519–2523. doi:10.1016/j.chb.2013.05.028

* Choi, B., Lee, I., Choi, D., & Kim, J. (2007). Collaborate and share: An experimental study of the effects of task and reward interdependencies in online games. *CyberPsychology & Behavior*, *10*(4), 591–595. doi:10.1089/cpb.2007.9985

Choi, Y. K., & Lee, J.-G. (2012). The persuasive effects of character presence and product type on responses to advergames. *Cyberpsychology, Behavior and Social Networking*, *15*(9), 503–506. doi:10.1089/cyber.2012.0012

* Christopherson, K. M., & Weatherly, J. N. (2006). The effect of visual complexity when playing a slot-machine simulation: The role of computer experience, computer anxiety, and optimism. *Computers in Human Behavior*, *22*(6), 1072–1079. doi:10.1016/j.chb.2004.03.031

Chua, P.-H., Jung, Y., Lwin, M. O., & Theng, Y.-L. (2013). Let's play together: Effects of video-game play on intergenerational perceptions among youth and elderly participants. *Computers in Human Behavior*, *29*(6), 2303–2311. doi:10.1016/j.chb.2013.04.037

Chumbley, J., & Griffiths, M. (2006). Affect and the computer game player: The effect of gender, personality, and game reinforcement structure on affective responses to computer game-play. *CyberPsychology & Behavior*, *9*(3), 308–316. doi:10.1089/cpb.2006.9.308

Ciavarro, C., Dobson, M., & Goodman, D. (2008). Implicit learning as a design strategy for learning games: Alert Hockey. *Computers in Human Behavior*, *24*(6), 2862–2872. doi:10.1016/j.chb.2008.04.011

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*(5), 587–591. doi:10.1089/cyber.2009.0263

* De Simone, J. J., Verbruggen, T., Kuo, L.-H., & Mutlu, B. (2012). Is cheating a human function? The roles of presence, state hostility, and enjoyment in an unfair video game. *Computers in Human Behavior*, *28*(6), 2351–2358. doi:10.1016/j.chb.2012.07.005

DeLeeuw, K. E., & Mayer, R. E. (2011). Cognitive consequences of making computer-based learning activities more game-like. *Computers in Human Behavior*, *27*(5), 2011–2016. doi:10.1016/j.chb.2011.05.008

* Eastin, M. S. (2006). Video game violence and the female game player: Self- and opponent gender effects on presence and aggressive thoughts. *Human Communication Research*, *32*(3), 351–372. doi:10.1111/j.1468-2958.2006.00279.x

Eastin, M. S. (2007). The influence of competitive and cooperative group game play on state hostility. *Human Communication Research*, *33*(4), 450–466. doi:10.1111/j.1468-2958.2007.00307.x

* Eastin, M. S., Appiah, O., & Cicchirllo, V. (2009). Identification and the influence of cultural stereotyping on postvideogame play hostility. *Human Communication Research*, *35*(3), 337–356. doi:10.1111/j.1468-2958.2009.01354.x

Eastin, M. S., & Griffiths, R. P. (2006). Beyond the shooter game: Examining presence and hostile outcomes among male game players. *Communication Research*, *33*(6), 448–466. doi:10.1177/0093650206293249

Echeverría, A., Améstica, M., Gil, F., Nussbaum, M., Barrios, E., & Leclerc, S. (2012). Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior*, *28*(4), 1170–1177. doi:10.1016/j.chb.2012.01.027

Ewoldsen, D. R., Eno, C. A., Okdie, B. M., Velez, J. A., Guadagno, R. E., & DeCoster, J. (2012). Effect of playing violent video games cooperatively or competitively on subsequent cooperative behavior. *Cyberpsychology, Behavior and Social Networking*, *15*(5), 277–280. doi:10.1089/cyber.2011.0308

Farrar, K. M., Krcmar, M., & Nowak, K. L. (2006). Contextual features of violent video games, mental models, andsAggression. *Journal of Communication*, *56*(2), 387–405. doi:10.1111/j.1460-2466.2006.00025.x

* Frey, A., Hartig, J., Ketzel, A., Zinkernagel, A., & Moosbrugger, H. (2007). The use of virtual environments based on a modification of the computer game Quake III Arena® in psychological experimenting. *Computers in Human Behavior*, *23*(4), 2026–2039. doi:10.1016/j.chb.2006.02.010

Gold, J. I., Kim, S. H., Kant, A. J., Joseph, M. H., & Rizzo, A. S. (2006). Effectiveness of virtual reality for pediatric pain distraction during IV placement. *CyberPsychology & Behavior*, *9*(2), 207–212. doi:10.1089/cpb.2006.9.207

Gordon, N. S., Merchant, J., Zanbaka, C., Hodges, L. F., & Goolkasian, P. (2011). Interactive gaming reduces experimental pain with or without a head mounted display. *Computers in Human Behavior*, *27*(6), 2123–2128. doi:10.1016/j.chb.2011.06.006

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: The virtual hospital experience. *Cyberpsychology, Behavior and Social Networking*, *14*(3), 99–105. doi:10.1089/cyber.2010.0100

Graham, J., Zheng, L., & Gonzalez, C. (2006). A cognitive approach to game usability and design: Mental model development in novice real-time strategy gamers. *CyberPsychology & Behavior*, *9*(3), 361–366. doi:10.1089/cpb.2006.9.361

Greitemeyer, T., Traut-Mattausch, E., & Osswald, S. (2012). How to ameliorate negative effects of violent video games on cooperation: Play it cooperatively in a team. *Computers in Human Behavior*, *28*(4), 1465–1470. doi:10.1016/j.chb.2012.03.009

Gross, M. L. (2010). Advergames and the effects of game-product congruity. *Computers in Human Behavior*, *26*(6), 1259–1265. doi:10.1016/j.chb.2010.03.034

Hämäläinen, R., Oksanen, K., & Häkkinen, P. (2008). Designing and analyzing collaboration in a scripted game for vocational education. *Computers in Human Behavior*, *24*(6), 2496–2506. doi:10.1016/j.chb.2008.03.010

* Hartmann, T., Toz, E., & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, *13*(4), 339–363. doi:10.1080/15213269.2010.524912

* Hartmann, T., & Vorderer, P. (2010). It's okay to shoot a character: Moral disengagement in violent video Games. *Journal of Communication*, *60*(1), 94–119. doi:10.1111/j.1460-2466.2009.01459.x

Hong, J.-C., & Liu, M.-C. (2003). A study on thinking strategy between experts and novices of computer games. *Computers in Human Behavior*, *19*(2), 245–258. doi:10.1016/S0747-5632(02)00013-4

Hou, J., Nam, Y., Peng, W., & Lee, K. M. (2012). Effects of screen size, viewing angle, and players' immersion tendencies on game experience. *Computers in Human Behavior*, *28*(2), 617–623. doi:10.1016/j.chb.2011.11.007

Hsu, S. H., Lee, F.-L., & Wu, M.-C. (2005). Designing action games for appealing to buyers. *CyberPsychology & Behavior*, *8*(6), 585–591. doi:10.1089/cpb.2005.8.585

Hsu, S. H., Wen, M.-H., & Wu, M.-C. (2007). Exploring design features for enhancing players' challenge in strategy games. *CyberPsychology & Behavior*, *10*(3), 393–397. doi:10.1089/cpb.2006.9940

Huang, W.-H. (2011). Evaluating learners' motivational and cognitive processing in an online game-based learning environment. *Computers in Human Behavior*, *27*(2), 694–704. doi:10.1016/j.chb.2010.07.021

Hummer, T. A., Wang, Y., Kronenberger, W. G., Mosier, K. M., Kalnin, A. J., Dunn, D. W., & Mathews, V. P. (2010). Short-term violent video game play by adolescents alters prefrontal activity during cognitive inhibition. *Media Psychology*, *13*(2), 136–154. doi:10.1080/15213261003799854

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*(3), 532–555. doi:10.1111/j.1460-2466.2007.00356.x

Ivory, J. D., & Magee, R. G. (2009). You can't take it with you? Effects of handheld portable media consoles on physiological and psychological responses to video game and movie content. *CyberPsychology & Behavior*, *12*(3), 291–297. doi:10.1089/cpb.2008.0279

Jannink, M. J. A., van der Wilden, G. J., Navis, D. W., Visser, G., Gussinklo, J., & Ijzerman, M. (2008). A low-cost video game applied for training of upper extremity function in children with cerebral palsy: A pilot study. *CyberPsychology & Behavior*, *11*(1), 27–32. doi:10.1089/cpb.2007.0014

* Jeong, E. J., & Biocca, F. A. (2012). Are there optimal levels of arousal to memory? Effects of arousal, centrality, and familiarity on brand memory in video games. *Computers in Human Behavior*, *28*(2), 285–291. doi:10.1016/j.chb.2011.09.011

* Jeong, E. J., Biocca, F. A., & Bohil, C. J. (2012). Sensory realism and mediated aggression in video games. *Computers in Human Behavior*, *28*(5), 1840–1848. doi:10.1016/j.chb.2012.05.002

Jerabeck, J. M., & Ferguson, C. J. (2013). The influence of solitary and cooperative violent video game play on aggressive and prosocial behavior. *Computers in Human Behavior*, *29*(6), 2573–2578. doi:10.1016/j.chb.2013.06.034

Jin, S.-A. A. (2009). Avatars mirroring the actual self versus projecting the ideal self: The effects of self-priming on interactivity and immersion in an exergame, Wii Fit. *CyberPsychology & Behavior*, *12*(6), 761–765. doi:10.1089/cpb.2009.0130

Jin, S.-A. A. (2010a). Does imposing a goal always improve exercise Intentions in avatar-based exergames? The moderating role of interdependent self-construal on exercise intentions and self-presence. *Cyberpsychology, Behavior and Social Networking*, *13*(3), 335–339. doi:10.1089/cyber.2009.0186

Jin, S.-A. A. (2010b). “I can be happy even when I lose the game”: The influence of chronic regulatory focus and primed self-construal on exergamers’ mood. *Cyberpsychology, Behavior and Social Networking*, *13*(4), 467–471. doi:10.1089/cyber.2009.0268

Jin, S.-A. A. (2010c). “I feel more connected to the physically ideal mini me than the mirror-image mini me”: Theoretical implications of the “malleable self” for speculations on the effects of avatar creation on avatar-self connection in Wii. *Cyberpsychology, Behavior and Social Networking*, *13*(5), 567–570. doi:10.1089/cyber.2009.0243

Jin, S.-A. A., & Park, N. (2009). Parasocial interaction with my avatar: Effects of interdependent self-construal and the mediating role of self-presence in an avatar-based console game, Wii. *CyberPsychology & Behavior*, *12*(6), 723–727. doi:10.1089/cpb.2008.0289

* Joeckel, S., Bowman, N. D., & Dogruel, L. (2012). Gut or game? The influence of moral intuitions on decisions in video games. *Media Psychology*, *15*(4), 460–485. doi:10.1080/15213269.2012.727218

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior*, *26*(6), 1246–1252. doi:10.1016/j.chb.2010.03.025

Jon Hobbs, L., & Yan, Z. (2008). Cracking the walnut: Using a computer game to impact cognition, emotion, and behavior of highly aggressive fifth grade students. *Computers in Human Behavior*, *24*(2), 421–438. doi:10.1016/j.chb.2007.01.031

Kim, K. J., & Sundar, S. S. (2013). Can interface features affect aggression resulting from violent video game play? An examination of realistic controller and large screen size. *Cyberpsychology, Behavior and Social Networking*, *16*(5), 329–334. doi:10.1089/cyber.2012.0500

* Kirsh, S. J., Olczak, P. V., & Mounts, J. R. W. (2005). Violent video games induce an affect processing bias. *Media Psychology*, *7*(3), 239–250. doi:10.1207/S1532785XMEP0801_4

* Klimmt, C., Hartmann, T., & Frey, A. (2007). Effectance and control as determinants of video game enjoyment. *CyberPsychology & Behavior*, *10*(6), 845–847. doi:10.1089/cpb.2007.9942

Klimmt, C., Hefner, D., Vorderer, P., Roth, C., & Blake, C. (2010). Identification with video game characters as automatic shift of self-perceptions. *Media Psychology*, *13*(4), 323–338. doi:10.1080/15213269.2010.524911

* Klimmt, C., Rizzo, A. S., Vorderer, P., Koch, J., & Fischer, T. (2009). Experimental evidence for suspense as determinant of video game enjoyment. *CyberPsychology & Behavior*, *12*(1), 29–31. doi:10.1089/cpb.2008.0060

* Knez, I., & Niedenthal, S. (2008). Lighting in digital game worlds: Effects on affect and play performance. *CyberPsychology & Behavior*, *11*(2), 129–137. doi:10.1089/cpb.2007.0006

Kohler, T., Fueller, J., Stieger, D., & Matzler, K. (2011). Avatar-based innovation: Consequences of the virtual co-creation experience. *Computers in Human Behavior*, *27*(1), 160–168. doi:10.1016/j.chb.2010.07.019

* Kozlov, M. D., & Johansen, M. K. (2010). Real behavior in virtual environments: Psychology experiments in a simple virtual-reality paradigm using video games. *Cyberpsychology, Behavior and Social Networking*, *13*(6), 711–714. doi:10.1089/cyber.2009.0310

Krcmar, M., Farrar, K. M., & McGloin, R. (2011). The effects of video game realism on attention, retention and aggressive outcomes. *Computers in Human Behavior*, *27*(1), 432–439. doi:10.1016/j.chb.2010.09.005

Krcmar, M., & Lachlan, K. (2009). Aggressive outcomes and videogame play: The role of length of play and the mechanisms at work. *Media Psychology*, *12*(3), 249–267. doi:10.1080/15213260903052257

Ku, J., Jang, H. J., Kim, K. U., Park, S. H., Kim, J. J., Kim, C. H., … Kim, S. I. (2006). Pilot study for assessing the behaviors of patients with schizophrenia towards a virtual avatar. *CyberPsychology & Behavior*, *9*(5), 531–539. doi:10.1089/cpb.2006.9.531

Lee, E.-J. (2004). Effects of gendered character representation on person perception and informational social influence in computer-mediated communication. *Computers in Human Behavior*, *20*(6), 779–799. doi:10.1016/j.chb.2003.11.005

Lee, E.-J. (2005a). When placebic information differs from real information: Cognitive and motivational bases of mindful reactions to informational social influence. *Communication Research*, *32*(5), 615–645. doi:10.1177/0093650205279351

Lee, E.-J. (2005b). Effects of the influence agent’s sex and self-confidence on informational social influence in computer-mediated communication: Quantitative versus verbal presentation. *Communication Research*, *32*(1), 29–58. doi:10.1177/0093650204271398

Lee, E.-J. (2007a). Character-based team identification and referent informational influence in computer-mediated communication. *Media Psychology*, *9*(1), 135–155. doi:10.1080/15213260709336806

Lee, E.-J. (2007b). Categorical person perception in computer-mediated communication: Effects of character representation and knowledge bias on sex inference and informational social influence. *Media Psychology*, *9*(2), 309–329. doi:10.1080/15213260701286007

Lee, E.-J. (2007c). Wired for gender: Experientiality and gender-stereotyping in computer-mediated communication. *Media Psychology*, *10*(2), 182–210. doi:10.1080/15213260701375595

Lee, K. M., Peng, W., & Klein, J. (2010). Will the experience of playing a violent role in a video game influence people's judgments of violent crimes? *Computers in Human Behavior*, *26*(5), 1019–1023. doi:10.1016/j.chb.2010.03.002

Lim, S., & Lee, J.-E. R. (2009). When playing together feels different: Effects of task types and social contexts on physiological arousal in multiplayer online gaming contexts. *CyberPsychology & Behavior*, *12*(1), 59–61. doi:10.1089/cpb.2008.0054

Lim, S., & Reeves, B. (2009). Being in the game: Effects of avatar choice and point of view on psychophysiological responses during play. *Media Psychology*, *12*(4), 348–370. doi:10.1080/15213260903287242

Limperos, A. M., Schmierbach, M., Kegerise, A. D., & Dardis, F. E. (2011). Gaming across different consoles: Exploring the influence of control scheme on game-player enjoyment. *Cyberpsychology, Behavior and Social Networking*, *14*(6), 345–350. doi:10.1089/cyber.2010.0146

Lin, J.-H. (2013a). Do video games exert stronger effects on aggression than film? The role of media interactivity and identification on the association of violent content and aggressive outcomes. *Computers in Human Behavior*, *29*(3), 535–543. doi:10.1016/j.chb.2012.11.001

Lin, J.-H. (2013b). Identification matters: A moderated mediation model of media interactivity, character identification, and video game violence on aggression. *Journal of Communication*, *63*(4), 682–702. doi:10.1111/jcom.12044

Lin, S.-F. (2010). Gender differences and the effect of contextual features on game enjoyment and responses. *Cyberpsychology, Behavior and Social Networking*, *13*(5), 533–537. doi:10.1089/cyber.2009.0293

Lin, S.-F. (2011). Effect of opponent type on moral emotions and responses to video game play. *Cyberpsychology, Behavior and Social Networking*, *14*(11), 695–698. doi:10.1089/cyber.2010.0523

* Malbos, E., Mestre, D. R., Note, I. D., & Gellato, C. (2008). Virtual reality and claustrophobia: Multiple components therapy involving game editor virtual environments exposure. *CyberPsychology & Behavior*, *11*(6), 695–697. doi:10.1089/cpb.2007.0246

Markey, P. M., & Scherer, K. (2009). An examination of psychoticism and motion capture controls as moderators of the effects of violent video games. *Computers in Human Behavior*, *25*(2), 407–411. doi:10.1016/j.chb.2008.10.001

McGloin, R., Farrar, K. M., & Krcmar, M. (2013). Video games, immersion, and cognitive aggression: Does the controller matter? *Media Psychology*, *16*(1), 65–87. doi:10.1080/15213269.2012.752428

* Michaliszyn, D., Marchand, A., Bouchard, S., Martel, M.-O., & Poirier-Bisson, J. (2010). A randomized, controlled clinical trial of in virtuo and in vivo exposure for spider phobia. *Cyberpsychology, Behavior and Social Networking*, *13*(6), 689–695. doi:10.1089/cyber.2009.0277

Nacke, L. E., Nacke, A., & Lindley, C. A. (2009). Brain training for silver gamers: Effects of age and game form on effectiveness, efficiency, self-assessment, and gameplay experience. *CyberPsychology & Behavior*, *12*(5), 493–499. doi:10.1089/cpb.2009.0013

Noyes, J. M., & Garland, K. J. (2003). Solving the Tower of Hanoi: does mode of presentation matter? *Computers in Human Behavior*, *19*(5), 579–592. doi:10.1016/S0747-5632(03)00002-5

Okazaki, S., & Yagüe, M. J. (2012). Responses to an advergaming campaign on a mobile social networking site: An initial research report. *Computers in Human Behavior*, *28*(1), 78–86. doi:10.1016/j.chb.2011.08.013

Okita, S. Y. (2004). Effects of age on associating virtual and embodied toys. *CyberPsychology & Behavior*, *7*(4), 464–471. doi:10.1089/cpb.2004.7.464

Orun, A., & Seker, H. (2012). Development of a computer game-based framework for cognitive behaviour identification by using Bayesian inference methods. *Computers in Human Behavior*, *28*(4), 1332–1341. doi:10.1016/j.chb.2012.02.017

* Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior, 24*(5), 2415–2433. doi:10.1016/j.chb.2008.02.016

Peng, W., & Crouse, J. (2013). Playing in parallel: The effects of multiplayer modes in active video game on motivation and physical exertion. *Cyberpsychology, Behavior and Social Networking*, *16*(6), 423–427. doi:10.1089/cyber.2012.0384

Peng, W., & Hsieh, G. (2012). The influence of competition, cooperation, and player relationship in a motor performance centered computer game. *Computers in Human Behavior*, *28*(6), 2100–2106. doi:10.1016/j.chb.2012.06.014

Peng, W., Lee, M., & Heeter, C. (2010). The effects of a serious game on role-taking and willingness to help. *Journal of Communication*, *60*(4), 723–742. doi:10.1111/j.1460-2466.2010.01511.x

Peng, W., Lin, J.-H., Pfeiffer, K. A., & Winn, B. (2012). Need satisfaction supportive game features as motivational determinants: An experimental study of a self-determination theory guided exergame. *Media Psychology*, *15*(2), 175–196. doi:10.1080/15213269.2012.673850

Peng, W., Liu, M., & Mou, Y. (2008). Do aggressive people play violent computer games in a more aggressive way? Individual difference and idiosyncratic game-playing experience. *CyberPsychology & Behavior*, *11*(2), 157–161. doi:10.1089/cpb.2007.0026

Persky, S., & Blascovich, J. (2007). Immersive virtual environments versus traditional platforms: Effects of violent and nonviolent video game play. *Media Psychology*, *10*(135-156). doi:10.1080/15213260701301236

Pfister, R. (2011). Gender effects in gaming research: a case for regression residuals? *Cyberpsychology, Behavior and Social Networking*, *14*(10), 603–606. doi:10.1089/cyber.2010.0547

Poels, K., van den Hoogen, W., Ijsselsteijn, W., & de Kort, Y. (2012). Pleasure to play, arousal to stay: The effect of player emotions on digital game preferences and playing time. *Cyberpsychology, Behavior and Social Networking*, *15*(1), 1–6. doi:10.1089/cyber.2010.0040

Ravaja, N. (2009). The psychophysiology of digital gaming: The effect of a non co-located opponent. *Media Psychology*, *12*(3), 268–294. doi:10.1080/15213260903052240

Ravaja, N., Saari, T., Salminen, M., Laarni, J., & Kallinen, K. (2006). Phasic emotional reactions to video game events: A psychophysiological investigation. *Media Psychology*, *8*(4), 343–367. doi:10.1207/s1532785xmep0804_2

Reinecke, L., Klatt, J., & Krämer, N. C. (2011). Entertaining media use and the satisfaction of recovery needs: Recovery outcomes associated with the use of interactive and noninteractive entertaining media. *Media Psychology*, *14*(2), 192–215. doi:10.1080/15213269.2011.573466

Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W. L. (2009). Multimodality and interactivity: Connecting properties of serious games with educational outcomes. *CyberPsychology & Behavior*, *12*(6), 691–697. doi:10.1089/cpb.2009.0099

* Robillard, G., Bouchard, S., Fournier, T., & Renaud, P. (2003). Anxiety and presence during VR immersion: A comparative study of the reactions of phobic and non-phobic participants in therapeutic virtual environments derived from computer games. *CyberPsychology & Behavior*, *6*(5), 467–476. doi:10.1089/109493103769710497

Roth, C., Vermeulen, I., Vorderer, P., Klimmt, C., Pizzi, D., Lugrin, J.-L., & Cavazza, M. (2012). Playing in or out of character: user role differences in the experience of interactive storytelling. *Cyberpsychology, Behavior and Social Networking*, *15*(11), 630–633. doi:10.1089/cyber.2011.0621

Salminen, M., & Ravaja, N. (2007). Oscillatory brain responses evoked by video game events: The case of super monkey ball 2. *CyberPsychology & Behavior*, *10*(3), 330–338. doi:10.1089/cpb.2006.9947

Schmierbach, M. (2010). “Killing spree”: Exploring the connection between competitive game play and aggressive cognition. *Communication Research*, *37*(2), 256–274. doi:10.1177/0093650209356394

Schmierbach, M., Limperos, A. M., & Woolley, J. K. (2012). Feeling the need for (personalized) speed: How natural controls and customization contribute to enjoyment of a racing game through enhanced immersion. *Cyberpsychology, Behavior and Social Networking*, *15*(7), 364–369. doi:10.1089/cyber.2012.0025

Schmierbach, M., Xu, Q., Oeldorf-Hirsch, A., & Dardis, F. E. (2012). Electronic friend or virtual foe: Exploring the role of competitive and cooperative multiplayer video game modes in fostering enjoyment. *Media Psychology*, *15*(3), 356–371. doi:10.1080/15213269.2012.702603

Schneider, E. F. (2004). Death with a story: How story impacts emotional, motivational, and physiological responses to First-Person Shooter video games. *Human Communication Research*, *30*(3), 361–375. doi:10.1111/j.1468-2958.2004.tb00736.x

* Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*, *28*(2), 648–658. doi:10.1016/j.chb.2011.11.011

Shafer, D. M. (2012). Causes of state hostility and enjoyment in player versus player and player versus environment video games. *Journal of Communication*, *62*(4), 719–737. doi:10.1111/j.1460-2466.2012.01654.x

Smith, S. P., Stibric, M., & Smithson, D. (2013). Exploring the effectiveness of commercial and custom-built games for cognitive training. *Computers in Human Behavior*, *29*(6), 2388–2393. doi:10.1016/j.chb.2013.05.014

Smyth, J. M. (2007). Beyond self-selection in video game play: An experimental examination of the consequences of massively multiplayer online role-playing game play. *CyberPsychology & Behavior, 10*(5), 717–721. doi:10.1089/cpb.2007.9963

Song, H., Kim, J., Kwon, R. J., & Jung, Y. (2013). Anti-smoking educational game using avatars as visualized possible selves. *Computers in Human Behavior, 29*(5), 2029–2036. doi:10.1016/j.chb.2013.04.008

Song, H., Kim, J., Tenzek, K. E., & Lee, K. M. (2013). The effects of competition and competitiveness upon intrinsic motivation in exergames. *Computers in Human Behavior*, *29*(4), 1702–1708. doi:10.1016/j.chb.2013.01.042

Song, S., & Kim, M. (2012). Five models of players' rule behavior for game balance. *Cyberpsychology, Behavior and Social Networking*, *15*(9), 498–502. doi:10.1089/cyber.2011.0504

Stouten, H., Heene, A., Gellynck, X., & Polet, H. (2012). Learning from playing with microworlds in policy making: An experimental evaluation in fisheries management. *Computers in Human Behavior*, *28*(2), 757–770. doi:10.1016/j.chb.2011.12.002

Takatalo, J., Kawai, T., Kaistinen, J., Nyman, G., & Häkkinen, J. (2011). User experience in 3D stereoscopic games. *Media Psychology*, *14*(4), 387–414. doi:10.1080/15213269.2011.620538

Tanes, Z., & Cho, H. (2013). Goal setting outcomes: Examining the role of goal interaction in influencing the experience and learning outcomes of video game play for earthquake preparedness. *Computers in Human Behavior*, *29*(3), 858–869. doi:10.1016/j.chb.2012.11.003

Teng, S. K. Z., Chong, G. Y. M., Siew, A. S. C., & Skoric, M. M. (2011). Grand Theft Auto IV comes to Singapore: Effects of repeated exposure to violent video games on aggression. *Cyberpsychology, Behavior and Social Networking*, *14*(10), 597–602. doi:10.1089/cyber.2010.0115

Tobin, S., & Grondin, S. (2009). Video games and the perception of very long durations by adolescents. *Computers in Human Behavior*, *25*(2), 554–559. doi:10.1016/j.chb.2008.12.002

Trepte, S., & Reinecke, L. (2011). The pleasures of success: Game-related efficacy experiences as a mediator between player performance and game enjoyment. *Cyberpsychology, Behavior and Social Networking*, *14*(9), 555–557. doi:10.1089/cyber.2010.0358

Tzeng, J.-Y., & Chen, C.-T. (2012). Computer praise, attributional orientations, and games: A reexamination of the CASA theory relative to children. *Computers in Human Behavior*, *28*(6), 2420–2430. doi:10.1016/j.chb.2012.07.013

Valadez, J. J., & Ferguson, C. J. (2012). Just a game after all: Violent video game exposure and time spent playing effects on hostile feelings, depression, and visuospatial cognition. *Computers in Human Behavior*, *28*(2), 608–616. doi:10.1016/j.chb.2011.11.006

* Valtchanov, D., Barton, K. R., & Ellard, C. (2010). Restorative effects of virtual nature settings. *Cyberpsychology, Behavior and Social Networking*, *13*(5), 503–512. doi:10.1089/cyber.2009.0308

* Van den Hoogen, W., Poels, K., IJsselsteijn, W., & de Kort, Y. (2012). Between challenge and defeat: Repeated player-death and game enjoyment. *Media Psychology*, *15*(4), 443–459. doi:10.1080/15213269.2012.723117

Van der Meij, H., Leemkuil, H., & Li, J.-L. (2013). Does individual or collaborative self-debriefing better enhance learning from games? *Computers in Human Behavior*, *29*(6), 2471–2479. doi:10.1016/j.chb.2013.06.001

Van Schaik, P., Blake, J., Pernet, F., Spears, I., & Fencott, C. (2008). Virtual augmented exercise gaming for older adults. *CyberPsychology & Behavior*, *11*(1), 103–106. doi:10.1089/cpb.2007.9925

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, *29*(6), 2568–2572. doi:10.1016/j.chb.2013.06.033

Walshe, D. G., Lewis, E. J., Kim, S. I., O'Sullivan, K., & Wiederhold, B. K. (2003). Exploring the use of computer games and virtual reality in exposure therapy for fear of driving following a motor vehicle accident. *CyberPsychology & Behavior*, *6*(3), 329–334. doi:10.1089/109493103322011641

Walshe, D. G., Lewis, E. J., O'Sullivan, K., & Kim, S. I. (2005). Virtually driving: Are the driving environments "real enough" for exposure therapy with accident victims? An explorative study. *CyberPsychology & Behavior*, *8*(6), 532–537. doi:10.1089/cpb.2005.8.532

Weaver, A. J., & Lewis, N. (2012). Mirrored morality: An exploration of moral choice in video games. *Cyberpsychology, Behavior and Social Networking*, *15*(11), 610–614. doi:10.1089/cyber.2012.0235

Weber, R., Ritterfeld, U., & Mathiak, K. (2006). Does playing violent video games induce aggression? Empirical evidence of a functional magnetic resonance imaging study. *Media Psychology*, *8*(1), 39–60. doi:10.1207/S1532785XMEP0801_4

* Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., & Groner, R. (2008). Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, *24*(5), 2274–2291. doi:10.1016/j.chb.2007.11.002

Whitlock, L. A., McLaughlin, A. C., & Allaire, J. C. (2012). Individual differences in response to cognitive training: Using a multi-modal, attentionally demanding game-based intervention for older adults. *Computers in Human Behavior*, *28*(4), 1091–1096. doi:10.1016/j.chb.2012.01.012

Williams, D. (2006). Virtual cultivation: Online worlds, offline perceptions. *Journal of Communication*, *56*(1), 69–87. doi:10.1111/j.1460-2466.2006.00004.x

* Wirth, W., Hofer, M., & Schramm, H. (2012). The role of emotional involvement and trait absorption in the formation of spatial presence. *Media Psychology*, *15*(1), 19–43. doi:10.1080/15213269.2011.648536

Wirth, W., Ryffel, F., von Pape, T., & Karnowski, V. (2013). The development of video game enjoyment in a role playing game. *Cyberpsychology, Behavior and Social Networking*, *16*(4), 260–264. doi:10.1089/cyber.2012.0159

Wohn, D. Y., & Wash, R. (2013). A virtual "Room" with a cue: Detecting personality through spatial customization in a city simulation game. *Computers in Human Behavior, 29*(1), 155–159. doi:10.1016/j.chb.2012.07.029

Yang, H.-L., & Wang, C.-S. (2008). Product placement of computer games in cyberspace. *CyberPsychology & Behavior, 11*(4), 399–404. doi:10.1089/cpb.2007.0099

Yoo, S.-C., & Peña, J. (2011). Do violent video games impair the effectiveness of in-game advertisements? The impact of gaming environment on brand recall, brand attitude, and purchase intention. *Cyberpsychology, Behavior and Social Networking, 14*(7-8), 439–446. doi:10.1089/cyber.2010.0031

Fight Fire with Rainbows: The Effects of Displayed Violence, Difficulty, and Performance in Digital Games on Affect, Aggression, and Physiological Arousal

Julia Kneer

Erasmus University Rotterdam

Malte Elson

University of Münster

Florian F. Knapp

University of Cologne

Author Note

Julia Kneer, Department of Communication, Erasmus University Rotterdam; Malte Elson, Department of Communication, University of Münster; Florian F. Knapp, Department of Psychology, University of Cologne.

Correspondence to: Julia Kneer, , 48143 Rotterdam, the Netherlands.

E-Mail: ; Tel:

Fight Fire with Rainbows: The Effects of Displayed Violence, Difficulty, and Performance in Digital Games on Affect, Aggression, and Physiological Arousal

Abstract

There is a large amount of variables that need to be taken into account when studying the effects of violent content in digital games. One of those variables is difficulty. In the current study participants played a modified first-person shooter in one of four different conditions, with either high or low difficulty and high or low violence. We assessed number of kills and number of deaths as game performance. Neither the difficulty nor the displayed violence had an effect on psychophysiological arousal during play, or post-game aggressive cognitions, emotions, and behavior. Number of deaths was found to be a positive predictor for positive affect, but only in the low difficulty condition. Number of kills was a positive predictor for positive affect and a negative predictor for negative affect. Thus, this study corroborates previous research indicating that violence in games does not substantially influence human behavior or experience, and other game characteristics deserve more attention in game effects studies.

Challenged by Rainbows: The Effects of displayed Violence, Difficulty, and Game-Performance on Aggression and Physiological Arousal

Digital games account for a large part of today's media landscape (Quandt, Breuer, Festl, & Scharkow, 2013). It is thus not surprising that researchers have studied the influence of digital games on their players for about thirty years. Similar to research on other media, it has primarily focused on the negative impact of gaming and especially on the link between violent digital games and human aggression. The discussion on whether or not this link exists and how exactly violent games could lead to increased aggression continues (see the recent dispute between Elson & Ferguson, 2014a, 2014b, and Bushman & Huesmann, 2014, Krahé, 2014, & Warburton, 2014). Researchers have started considering distinctive differences of digital games compared to other media – for example the interactivity of the medium – as these are important factors when studying their effects. Previous research on the influence of digital games on aggression and physiological arousal indicates that a mono-causal connection between violent contents and these outcomes might be too simplistic. There is evidence that the characteristics of the recipient of the game (Kneer, Munko, Glock, & Bente, 2012), the playing context (Breuer, Scharkow, & Quandt, 2013; Velez, Mahood, Ewoldsen, & Moyer-Guse, 2013), and game-specific characteristics (Elson, Breuer, Van Looy, Kneer, & Quandt, 2013) have to be taken into account.

Concerning game characteristics which might have an influence on aggression besides violent content, Adachi and Willoughby (2011b) name competition, pace of action, and difficulty as possible confounds. While there have been studies on the interaction of the first two factors with violent content (Adachi & Willoughby, 2011a; Elson et al., 2013), it remains unclear is how the third variable – *difficulty* of the game – is influencing results on online- and post-game outcomes. Therefore, the current study explores the effects of difficulty and violence of a digital game on human aggression and psychophysiological arousal. To test this we modified the displayed violent and difficulty of a first-person shooter, measured aggressive behaviors, aggression-related associations, physiological arousal, and positive and negative affect, while also accounting for the individual in-game performance.

Violent Digital Games and Aggression

Different models and theories have been developed to predict the effects of violent content in games on physiological arousal and aggressive cognitions, emotions, and behaviors. The General Aggression Model (GAM; Anderson & Bushman, 2002) integrates several older theories into one model and constitutes a social-cognitive approach to aggression. According to the GAM, aggressive behaviors are acquired through learning and reinforcement of knowledge-structures which influence the perception and interpretation of simple cues up to complex behavior-sequences. Over a period of time and repeats these knowledge-structures, which can also include certain scripts, can become automated. According to the GAM, repeated exposure to violent media facilitates hostile perception of situational variables and behavior of others, and increases the accessibility of an aggressive repertoire. In the long term, the GAM predicts that this will result in fundamental changes and can foster an aggressive personality. Other approaches, such as the Catalyst Model (Ferguson, Rueda, et al., 2008), focus on biological determinants as well as the social context of family and peer groups in which aggression is fostered. These biopsychosocial models consider games in the etiology of aggression to be an unimportant influence.

Empirical Evidence for Violent Game Effects

Research concerning negative effects of digital games is still contested for almost any relevant outcome. For psychophysiological arousal (heart rate, HR; skin conductance level, SCL; and blood pressure) as an online-measurement during game play, the findings are ambivalent. Some studies found that violent games increase psychophysiological responses (Barlett, Harris, & Bruey, 2008), while other studies could not reveal these effects (Anderson & Carnagey, 2009). Research on the activation of aggression-related cognitions after playing violent games show the same inconclusive pattern. Some studies found a higher accessibility of aggressive thoughts (Anderson & Dill, 2000) while others found that aggressive thoughts are suppressed by experienced players (Glock & Kneer, 2009). Besides these controversy results on psychophysiological arousal and cognition, the measurement of aggressive behaviors is still discussed among scholars regarding the operationalization, standardization, and validity of measures used in laboratory studies (Elson, Mohseni, Breuer, Scharkow, & Quandt, 2014; Ferguson & Savage, 2012; Tedeschi & Quigley, 1996). It might be due to these methodological shortcomings that, not surprisingly, results concerning aggressive post-game behaviors are ambivalent, too. Some scholars report findings of increases in aggressive behaviors after playing violent games (Anderson et al., 2004), while others find mixed evidence (Anderson & Dill, 2000), or no link at all (Adachi & Willoughby, 2011a). For an exhaustive review of the empirical literature on these and other variables, see Elson and Ferguson (2014b).

One main criticism of many studies investigating the effect of violent content on aggressive behaviour and physiological arousal is the lack of stimulus control. Usually, participants are assigned to play one of two different games, a violent or a nonviolent one. However, violent content is usually not the *only* dimension on which the games used in these studies differ (Adachi & Willoughby, 2011b). These other dimensions might pose confounds that must be controlled so they do not interfere with any effect that should be explained exclusively by the manipulation. For a study to qualify as an experiment, however, all conditions must be held constant except for one factor which is the independent variable. Without sufficient stimulus control, manipulating game contents as independent variables by using completely different games could violate fundamental assumptions of experiments as a scientific method (Elson & Quandt, in press). Järvelä, Ekman, Kivikangas, and Ravaja (2014) also stress the importance of manipulation of one game instead of using different games, which can be done, for instance, through in-game settings or game modifications (“modding”). Modding is a relatively novel method to generate stimulus materials which is only differ in the relevant variables, thus increasing internal validity without necessarily decreasing external validity. For example, Hartmann, Toz, and Brandon (2010) created two mods to assess the effects of (un)justified violence on feelings of guilt: Their participants were either playing UN soldiers attempting to shut down a torture camp, or paramilitary forces defending the camp and continuing the cruelty. In the study by Staude-Müller, Bliesener, and Luthman (2008) participants either played a regular first-person shooter, or a modification in which avatars were being frozen instead of killed.

Digital Games and Aggression: The Role of Difficulty

Despite its obvious importance for the experience of games (van den Hoogen, Poels, IJsselsteijn, & de Kort, 2012) and for potential positive and negative outcomes of playing (Adachi & Willoughby, 2011a), game difficulty as a source of aggression has not been investigated yet. Challenge, determined by game difficulty and player skill, is one major requirement for game enjoyment. A game too easy (e.g., for an experienced player) could result in boredom, while a game too demanding (particularly for beginners) could lead to frustration. According to van den Hoogen et al. (2012), the optimal experience of games (resulting in a higher enjoyment) exists in the perfect balance between challenge and defeat (see also the literature on the concept of flow in games, e.g. Cowley, Charles, Black, & Hickey, 2008).

In addition, considering the classic frustration-aggression hypothesis (Berkowitz, 1989; Dollard, Miller, Doob, Mowrer, & Sears, 1939), frustration could be an important confound in the research on game violence and aggression. Concerning digital games, difficulty as well as game-performance should influence frustration and therefore, online- and post-game variables. Despite its relevance for explaining aggressive behaviors in particular, frustration has not seen a lot of attention in digital games research. Some studies identified self-reports of frustration as important control variables (Anderson et al., 2004; Valadez & Ferguson, 2012; Velez et al., 2013) without specifically testing frustration-related hypotheses. Ivory and Kalyanaraman (2007) conclude that "research that intentionally manipulates frustration as an independent variable might prove insightful" (p. 551) when trying to understand media effects as more than just issues of content. The frustration-aggression hypothesis is mentioned explicitly as a potential explanation for the effects of (violent) digital games by several authors (Eastin & Griffiths, 2006; R. Williams & Clippinger, 2002). So far, only two studies tested this hypothesis for digital games: Schmierbach (2010) found no mediating effect of frustration on aggressive cognitions after violent game exposure, while Breuer et al. (2013) found that frustration mediated the effect of outcome (losing) in a non-violent game on aggressive behavior. This is further corroborated by the study of Shafer (2012), who found that undesired outcomes in competitive situation can facilitate hostility.

Assumptions and research questions

With these theoretical considerations in mind, we elicited the following hypotheses for our experiment:

H1) A larger amount of violent content increases a) physiological arousal during game-play, b) accessibility of aggressive thoughts, c) aggressive behavior, d) negative post-game emotions, and e) decreases positive post-game emotions compared to a lower amount of violent content.

H2) Higher game difficulty increases a) physiological arousal during game-play, b) accessibility of aggressive thoughts, c) aggressive behavior, d) negative post-game emotions, and e) decreases positive post-game emotions compared to a lower amount of violent content.

H3) In addition, we assumed an interaction of difficulty and violent content, leading to the highest values for a) physiological arousal during game-play, b) accessibility of aggressive thoughts, c) aggressive behavior, d) negative emotions and e) lowest positive emotions in the high difficulty / high violence condition compared to the other conditions.

RQ1) What impact has the number of kills on online measures and post-game outcomes?

RQ2) What impact has the number of deaths on online measures and post-game outcomes?

Methods

Design and Procedure

Participants (*N* = 90) played the first-person shooter *Team Fortress 2* (Valve, 2007) in which two teams both try to capture and hold a control point while preventing the other team from doing so (in this study, teammates and the opponent team were controlled by the computer). Participants played in one of four different conditions with either high or low difficulty and a high or low amount of violent content.

After entering the lab and signing an informed consent form, the participants were briefed about the upcoming experiment. Before the game was started, participants got an introduction to the lexical decision task (LDT) and the competitive reaction time task (CRTT) to save time after playing so that possible effects of the game would still last during the measurement of possible aggressive cognitions and behaviors (Anderson, Gentile, & Buckley, 2007). After the introduction electrodes were attached to the participants' fingers and baseline measures were taken. Participants were told that they would play against six computer-controlled opponents (bots). First, participants could accustom to the game, the map, and the controls in a free training round without enemies. This training round would go on until the participants said they mastered the controls. After the training round, the experimental session started and participants played for 20 minutes. Game performance was recorded through event log files, which were later analyzed using a log parsing program (Fausak, 2008) to calculate in-game-statistics: kills (how often the player killed an enemy) and deaths (how often the player avatar was killed).

After finishing the 20-minute session, participants had to complete the LDT and CRTT. Following the CRTT, the experimenter opened a browser window with a questionnaire to assess post-game emotions. At the end of the experiment participants were thanked and debriefed. The whole procedure lasted for about 45 minutes.

Stimulus Material

To generate stimulus materials for each experimental group, we created four different versions of *Team Fortress 2* instead of using four different games, thus allowing the exact altering of the independent variables without introducing potentially confounding factors. The easy or hard difficulty was achieved by changing the player's weapon's damage output, the player's resistance to enemy damage, and the speed at which the control point could be captured. Table 1 shows the values in each condition.

INCLUDE TABLE 1

In the high violent conditions, the player and all bots wielded flamethrowers, and the portrayed deaths of characters in the game were rather bloody and graphic. In the low violent conditions, everyone was equipped with a 'rainbowblower' that exhausted rainbows instead of fire while playing bubbling sounds, and instead of dying, this weapon incapacitated characters by making them drop to the ground convulsing with laughter (see figure 1 for screenshot examples).

INSERT FIGURE 1

Treatment Check

To verify whether the manipulations were successful or not, participants were asked to indicate the age rating they would recommend for game they just played. The participants could choose from *no age restriction*, restrictions for those *below the age of 6*, *12, 16*, *18,* and *no clearance* (taken from German rating system participants were likely to be familiar with). A Mann-Whitney U test revealed that participants in the high violent conditions chose a significantly higher age rating ($Mo = 18$) than participants in the non-violent conditions ($Mo = 12$), $U = 433.00$, $z = -4.28$, $p < .001$, $r = -.48$. In addition, we asked the participants about their perception of the game they played. One-way ANOVAs showed that the perception of the participant differed significant for the subjective perceived violence, $F(1,82) = 20.17$, $p < .001$, $\omega^2 = .18$, and for the perceived difficulty, $F(1,82) = 40.28$, $p < .001$, $\omega^2 = .32$.

INSERT TABLE 1 ABOUT HERE

Sample

Participants ($N = 90$; 21 males and 69 females) were undergraduate and graduate students from a large German university recruited via the online recruitment tool *Cortex* (Elson & Bente, 2009). Due to technical difficulties, data from six participants had to be discarded from further analysis, leaving a total of 84, with 21 participants in each of the four game conditions. Gender was equality distributed over all conditions, $\chi^2(3) = 1.58$, n.s. Distribution of experienced ($n = 58$) and unexperienced players ($n = 26$) was equal for all experimental conditions, $\chi^2(3) = 2.01$, n.s.. Experienced players reported to play digital games for $M = 4.04$ hours per week ($SD = 4.87$). Mean age of the participants was

M = 24.47 years (SD = 6.24).

Measurements

Psychophysiological arousal. To measure the participants' physiological arousal we recorded interbeat intervals (IBI) and electrodermal activity (EDA) with the Wild Divine IOM Lightstone Biometrics USB Widget. This device provides three plastic finger clips, two for SCL, and one for HR. The clips were attached to the player's thumb, ring finger, and middle or left finger (for all participants, as none used the left hand for the mouse). The IOM'S plastic cases and the mouse used to control the game were wrapped with hook-and-loop tape, sparing only the left mouse button. Data of both indicators were separated into six different time blocks in order to count for changes over time: base-line and four playing phases of five minutes each.

Cognition. To asses aggression-related associations, we used a lexical decision task similar to Kneer et al. (2012). Participants were presented with a series of 44 letter strings for each of which they had to indicate whether it was a proper word of the German language or if it was a non-word. Altogether there were 11 aggressive, 11 neutral, and 22 non-words. Shorter latencies for aggressive words indicate a higher accessibility of aggression-related association. The first four words were fixed in the presentation and later excluded from analysis to eliminate the influence of practice. The remaining 40 letter strings were presented in a randomized order for each participant. Data were checked for extreme outliers in response latencies and low accuracy ratios. Response latencies were then standardized with the length of the letter sting and mean latencies for aggressive words and neutral words were calculated.

Aggressive behavior. Aggressive behavior was measured with the standardized version of the Competitive Reaction Time Task (Ferguson, Smith, Miller-Stratton, Fritz, & Heinrich, 2008). Participants were told that they would play 25 rounds of a reaction time game against another person in an adjacent room. The task was to push a button as fast as possible after an optical signal. Before each of the 25 trials, participants have to set the volume and duration of a noise blast (on a scale from 1 to 10) their opponent will hear in case the opponent loses that round. They are told that their opponent will set the volume and duration of a noise blast as well. If the participant loses a round, he or she hears a noise blast with the settings allegedly chosen by the opponent. The settings chosen by their "opponent" are shown on screen after each trial. In reality, there is no other participant and the sequence of wins and losses, volume and duration settings are randomized and preset. The first trial is always a loss, and the opponent's settings are volume 5, duration 5. After that, there are 12 wins and 12 losses over 24 trials with volume and duration settings ranging from 2 to 9. We calculated the mean duration and mean intensity over all 25 trials as separate scores for aggressive behavior.

Post-game emotions. Emotional outcome of the game was assessed via the M-DAS affect scale (Renaud & Unz, 2006), which includes the following subscales: joy, surprise, interest, fun, contentment, love, dolefulness, anger, disgust, contempt, fear, shame, guilt, fascination, enchantment, and boredom. The subscale 'love' was excluded from further analyses due to a low Cronbach's alpha (α = .52). For all other subscales the values ranged from α = .65 to.89.

Results

To test our hypotheses, we conducted separate one-way ANOVAs for the LDT response latencies, and mean volume and duration scores; rANOVAS for interbeat interval and electrodermal activity scores; and MANOVAs for the post-game affect scores. Research questions RQ1 and RQ2 were tested by conducting regression analyses with number of kills, number of deaths, and violence as predictors for each dependent variable separately for the high difficulty and the low difficulty conditions.

Psychophysiological Arousal

There was no significant effect on interbeat interval by difficulty, $F(1,51) = 0.45$, $p > .05$, $r = .09$, violence, $F(1,51) = 1.00$, $p > .05$, $r = .14$, or their interaction, $F(1,51) = 0.183$, $p > .05$, $r = .05$. Neither was there a significant effect on the electrodermal activity for difficulty, $F(1,50) = 0.63$, $p > .05$, $r = .11$, violence, $F(1,50) = 0.04$, $p > .05$, $r = .03$, or the interaction, $F(1,50) = 0.64$, $p > .05$, $r = .11$. Regression analyses did not reveal any significant predictor or model for both measurements.

Aggression-related cognitions

There was no significant effect of violent content on response latencies, but there was a significant effect for the difficulty, revealing that in the high difficulty condition participants responded slower to aggressive words than in the low difficulty condition (see Table 2). Regression analyses did not reveal any significant predictor or model for the response latencies for neutral or aggressive words.

INSERT TABLE 2

Aggressive behavior

One participant had to be excluded from the analysis, because she aborted the testing before finishing it, thus leaving data from $N = 83$ participants. Neither difficulty, nor violence, nor their interaction had a significant effect on either mean volume or mean duration (see Table 3). Regression analyses did not reveal any significant predictor or model for mean volume or duration.

INSERT TABLE 3

Post-game emotions

The MANOVA for post-game emotions did not reveal any significant effects of violence, $V = 0.23$, $F(16,65) = 1.19$, $p > .05$, difficulty, $V = 0.23$, $F(16,65) = 1.23$, $p > .05$, or their interaction, $V = 0.20$, $F(16,65) = 1.00$, $p > .05$. In contrast to these results, we found significant results for the regression analyses. For the positive emotions fun, pleasure, satisfaction, and interest, number of kills was found to be a significant predictor in both difficulty conditions. In addition, number of deaths was a significant predictor for fun, pleasure, and satisfaction but only in the low difficulty condition resulting in a positive relationship (an increase in number of deaths resulted in an increase of those positive emotions). Concerning the negative emotions, disgust, contempt, and fear, number of kills was found to have a significant negative predictive value (the higher the number of kills the less participants experienced negative emotions) but only in the low difficulty condition (see Table 4). Number of deaths did not predict negative emotions.

INSERT TABLE 4

Discussion

In sum, we did not find that violent content had any effect on psychophysiological arousal, aggressive cognitions or behavior, and post-game affect, and thus, hypotheses H1a, b, c, d were not supported. This corroborates previous findings (Adachi & Willoughby, 2011a; Valadez & Ferguson, 2012), while it stands in opposition to others (Anderson & Carnagey, 2009; Anderson et al., 2004). Difficulty of a game did not influence arousal, behavior, and emotions. Thus, hypotheses H2a, c, d were not supported. There was a significant effect on the activation of aggressive associations through an increased difficulty but not in the assumed direction (H2b). This difference was not significant for neutral words but the trend was similar, showing higher response latencies when the difficulty was increased. The reason for this finding might be that a higher difficulty of a game leads to exhaustion, resulting in slower responses in the lexical decision task in general. As for arousal, behavior, and emotions, we again did not find any interaction between violence and difficulty. Number of kills and deaths did not influence the results of arousal, cognition, and behavior neither for the high nor for the low difficulty conditions.

Number of kills predicted both positive and negative post-game affect. This is not surprising, as success in a game should increase positive and decrease negative feelings. In addition, deaths stimulated positive emotions but only in the low difficulty condition. Arguably, the positive emotions elicited through success increase in particular when players feel stimulated and challenged (van den Hoogen et al., 2012). Without a minimal amount of challenge, games might simply be too boring, which should decrease enjoyment and result in lower positive affect. In contrast, number of deaths did not predict positive or negative feelings in the high difficulty condition. Therefore, frustration again did not influence emotions as outcome after playing a game, and it did not have an impact on measures of arousal and aggression

One reason for this finding might be that these measures do not capture frustration to its full extent. According to the frustration-aggression hypothesis (Berkowitz, 1989), frustration is defined as an event blocking a goal that is *expected* and *important* to the individual (see also the behavioral definition of frustration by Brown and Farber, 1951). However, we did not manipulate or assess these two requirements in our study. Given the large variability of game experience in our sample, it is plausible that there was also a large variability in the expected outcome of playing. While experienced players probably have a higher expectation to win (and they should be frustrated if they lose), inexperienced ones probably have no expectation at all towards their performance. Similarly, it cannot be said that the study design made it important to win, as there was no incentive based on performance. Again, personal characteristics in individuals, such as competitiveness, might increase the importance of winning for some, but not for all of our participants. Therefore, the emotions elicited by kills and deaths might not have been strong enough to increase aggressiveness. Future research should specifically manipulate expectations and importance of outcomes.

Limitations

There are several limitations of this study which need to be taken into account. First, while the low violence condition was not specifically created for this research purpose and had thus a high ecological validity, it is still possible that many players might find it a rather unusual way of playing. Particularly experienced players who are used to violent first-person shooters might have been bothered by this version, which could present a potential influence on the outcomes. Expertise of players is another important factor. Here it could prove interesting to study differences between novice and experienced players, which expectations they have regarding their own performance, and what happens if these expectations are not met.

Finally, there is a heated debate on issues of standardization, reliability, and external validity regarding our measure for aggressive behavior, the CRTT (Elson et al., 2014; Ritter & Eslea, 2005; Tedeschi & Quigley, 1996), which might be a further limitation of this study. To avoid the problem of lacking standardization, we used the standardized version suggested by Ferguson et al. (2008). However, this does not solve the issue of lacking evidence for the CRTT's external validity, and as such, our findings regarding aggressive behaviors have to be interpreted with due care.

Conclusion

This study focused on the influence of difficulty, violent content, and in-game success on arousal, aggression-related associations, aggressive behavior, and post-game emotions. Once again, there was no effect of violent content on aggression-related associations. Difficulty did only influence cognitive processes; probably due to exhaustion of participants in the high difficulty condition. Despite the zero effects of violent content and difficulty on post-game emotion, in-game success (number of kills) as well as challenge (number of deaths) were found to influence post-game emotional experience and, thus, should be considered as important influence factors in future game research.

References

Adachi, P. J. C., & Willoughby, T. (2011a). The effect of video game competition and violence on aggressive behavior: Which characteristic has the greatest influence? *Psychology of Violence, 1*(4), 259–274. doi:10.1037/a0024908

Adachi, P. J. C., & Willoughby, T. (2011b). The effect of violent video games on aggression: Is it more than just the violence? *Aggression and Violent Behavior, 16*(1), 55–62. doi:10.1016/j.avb.2010.12.002

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*(1), 27–51. doi:10.1146/annurev.psych.53.100901.135231

Anderson, C. A., & Carnagey, N. L. (2009). Causal effects of violent sports video games on aggression: Is it competitiveness or violent content? *Journal of Experimental Social Psychology*, *45*(4), 731–739. doi:10.1016/j.jesp.2009.04.019

Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin, A. J., Eubanks, J., & Valentine, J. C. (2004). Violent video games: Specific effects of violent content on aggressive thoughts and behavior. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 36, pp. 199–249). New York, NY: Elsevier. doi:10.1016/S0065-2601(04)36004-1

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, *78*(4), 772–790. doi:10.1037/0022-3514.78.4.772

Anderson, C. A., Gentile, D. A., & Buckley, K. E. (2007). *Violent video game effects on children and adolescents: Theory, research, and public policy*. New York, NY: Oxford University Press.

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*(3), 539–546. doi:10.1016/j.jesp.2007.10.003

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, *106*(1), 59–73. doi:10.1037/0033-2909.106.1.59

Breuer, J., Scharkow, M., & Quandt, T. (2013). Sore losers? A reexamination of the frustration-aggression hypothesis for colocated video game play. *Psychology of Popular Media Culture*. Advance online publication. doi:10.1037/ppm0000020

Brown, J. S., & Farber, I. E. (1951). Emotions conceptualized as intervening variables - with suggestions toward a theory of frustration. *Psychological Bulletin*, *48*(6), 465–495. doi:10.1037/h0058839

Bushman, B. J., & Huesmann, L. R. (2014). Twenty-five years of research on violence in digital games and aggression: A reply to Elson & Ferguson (2013). *European Psychologist*, *19*(1), 47–55. doi:10.1027/1016-9040/a000164

Cowley, B., Charles, D., Black, M., & Hickey, R. (2008). Toward an understanding of flow in video games. *Computers in Entertainment*, *6*(2). doi:10.1145/1371216.1371223

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. New Haven, CT: Yale University Press. doi:10.1037/10022-000

Eastin, M. S., & Griffiths, R. P. (2006). Beyond the shooter game: Examining presence and hostile outcomes among male game players. *Communication Research*, *33*(6), 448–466. doi:10.1177/0093650206293249

Elson, M., & Bente, G. (2009). CORTEX - Computer-Aided Registration Tool for Experiments. Cologne, Germany: University of Cologne. Retrieved from http://cortex.uni-koeln.de

Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (2013). Comparing apples and oranges? Evidence for pace of action as a confound in research on digital games and aggression. *Psychology of Popular Media Culture*. Advance online publication. doi:10.1037/ppm0000010

Elson, M., & Ferguson, C. J. (2014a). Does doing media violence research make one aggressive? The ideological rigidity of social cognitive theories of media violence and response to Bushman and Huesmann (2013), Krahé (2013), and Warburton (2013). *European Psychologist*, *19*(1), 68–75. doi:10.1027/1016-9040/a000185

Elson, M., & Ferguson, C. J. (2014b). Twenty-five years of research on violence in digital games and aggression: Empirical evidence, perspectives, and a debate gone astray. *European Psychologist, 19*(1), 33–46. doi:10.1027/1016-9040/a000147

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the Competitive Reaction Time Task in aggression research. *Psychological Assessment*. Advance online publication. doi:10.1037/a0035569

Elson, M., & Quandt, T. (in press). Digital games in laboratory experiments: Controlling a complex stimulus through modding. *Psychology of Popular Media Culture*. doi:10.1037/ppm0000033

Fausak, R. (2008). TF2 Log Parser (Version 1.1).

Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S., & Smith, S. M. (2008). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation? *Criminal Justice and Behavior*, *35*(3), 311–332. doi:10.1177/0093854807311719

Ferguson, C. J., & Savage, J. (2012). Have recent studies addressed methodological issues raised by five decades of television violence research? A critical review. *Aggression and Violent Behavior*, *17*(2), 129–139. doi:10.1016/j.avb.2011.11.001

Ferguson, C. J., Smith, S. M., Miller-Stratton, H., Fritz, S., & Heinrich, E. (2008). Aggression in the laboratory: Problems with the validity of the modified Taylor competitive reaction time test as a measure of aggression in media violence studies. *Journal of Aggression, Maltreatment & Trauma*, *17*(1), 118–132. doi:10.1080/10926770802250678

Glock, S., & Kneer, J. (2009). Game over? The impact of knowledge about violent digital games on the activation of aggression-related concepts. *Journal of Media Psychology: Theories, Methods, and Applications*, *21*(4), 151–160. doi:10.1027/1864-1105.21.4.151

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, *13*(4), 339–363. doi:10.1080/15213269.2010.524912

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*(3), 532–555. doi:10.1111/j.1460-2466.2007.00356.x

Järvelä, S., Ekman, I., Kivikangas, J. M., & Ravaja, N. (2014). A practical guide to using digital games as an experiment stimulus. *Transactions of the Digital Games Research Association*, *1*(2), 85–116.

Kneer, J., Munko, D., Glock, S., & Bente, G. (2012). Defending the doomed: Implicit strategies concerning protection of first-person shooter games. *Cyberpsychology, Behavior, and Social Networking*, *15*(5), 251–256. doi:10.1089/cyber.2011.0583

Krahé, B. (2014). Restoring the spirit of fair play in the debate about violent video games: A comment on Elson and Ferguson (2013). *European Psychologist*, *19*(1), 56–59. doi:10.1027/1016-9040/a000165

Quandt, T., Breuer, J., Festl, R., & Scharkow, M. (2013). Digitale Spiele: Stabile Nutzung in einem dynamischen Markt. [Digital games: Stable patterns of use in a dynamic market.]. *Media Perspektiven*, *10*, 483–492.

Renaud, D., & Unz, D. (2006). Die M-DAS - eine modifizierte Version der Differentiellen Affekt Skala zur Erfassung von Emotionen bei der Mediennutzung. *Zeitschrift Für Medienpsychologie*, *18*(2), 70–75. doi:10.1026/1617-6383.18.2.70

Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior*, *31*(5), 407–419. doi:10.1002/ab.20066

Schmierbach, M. (2010). “Killing spree”: Exploring the connection between competitive game play and aggressive cognition. *Communication Research*, *37*(2), 256–274. doi:10.1177/0093650209356394

Shafer, D. M. (2012). Causes of state hostility and enjoyment in player versus player and player versus environment video games. *Journal of Communication, 62*(4), 719–737. doi:10.1111/j.1460-2466.2012.01654.x

Staude-Müller, 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*(1), 41–50. doi:10.1024/1421-0185.67.1.41

Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior*, *1*(2), 163–177. doi:10.1016/1359-1789(95)00014-3

Valadez, J. J., & Ferguson, C. J. (2012). Just a game after all: Violent video game exposure and time spent playing effects on hostile feelings, depression, and visuospatial cognition. *Computers in Human Behavior*, *28*(2), 608–616. doi:10.1016/j.chb.2011.11.006

Valve. (2007). Team Fortress 2. Bellevue, WA: Valve.

Van den Hoogen, W. M., Poels, K., IJsselsteijn, W. A., & de Kort, Y. A. W. (2012). Between challenge and defeat: Repeated player-death and game enjoyment. *Media Psychology*, *15*(4), 443–459. doi:10.1080/15213269.2012.723117

Velez, J. A., Mahood, C., Ewoldsen, D. R., & Moyer-Guse, E. (2013). Ingroup versus outgroup conflict in the context of violent video game play: The effect of cooperation on increased helping and decreased aggression. *Communication Research*. doi:10.1177/0093650212456202

Warburton, W. (2014). Apples, oranges and the burden of proof - Putting media violence findings in context. A comment on Elson and Ferguson (2013). *European Psychologist*, *19*(1), 60–67. doi:10.1027/1016-9040/a000166

Williams, R. B., & Clippinger, C. A. (2002). Aggression, competition and computer games: Computer and human opponents. *Computers in Human Behavior*, *18*(5), 495–506. doi:10.1016/S0747-5632(02)00009-2

Table 1. Overview of the changes in the game stats concerning high and low difficulty.

	Difficulty	
	high	low
Weapon Damage [a]	75%	200%
Vulnerability towards enemy fire [a]	125%	75%
Control point capturing speed [a]	100%	500%

Note. [a] Compared to default value

Table 2. Results of ANOVAs for aggressive and neutral word types of the lexical decision task with *F*-values and effect sizes (ω^2) ($N = 84$).

	Violence		Difficulty		Violence x Difficulty	
	F (1,80)	ω^2	F (1,80)	ω^2	F (1,80)	ω^2
Neutral	0.00 n.s.	.00	3.15 n.s.	.03	0.03 n.s.	.00
Aggression	0.00 n.s.	.00	8.58 **	.08	0.46 n.s.	.00

Note. n.s. = not significant, ** $p < .01$

Table 3. Results of ANOVAs for average noise duration, average noise intensity, and high intensity aggression (HIA) with *F*-values and effect sizes (ω^2) (N = 83).

	Violence		Difficulty		Violence x Difficulty	
	F (1,79)	ω^2	F (1,79)	ω^2	F (1,79)	ω^2
Duration	0.42 n.s.	.00	0.15 n.s.	.00	0.15 n.s.	.00
Intensity	0.29 n.s.	.00	0.32 n.s.	.00	0.01 n.s.	.00

Note. n.s. = not significant.

Table 4. beta Weights and R^2 of the regression analyses with post-game emotions as criteria.

Post-game Emotion	predictor	High Difficulty	Low Difficulty
Fun	Kills	.36*	.65**
	Deaths	-.20	.34*
	Violence	-.09	-.13
R^2_{adj}		.15	.31***
Pleasure	Kills	.40*	.68***
	Deaths	-.22	.40**
	Violence	.01	-.17
R^2_{adj}		.16	.35***
Satisfaction	Kills	.39*	.41*
	Deaths	-.21	.18
	Violence	.07	-.03
R^2_{adj}		.14	.12
Interest	Kills	.41*	.24
	Deaths	-.14	.06
	Violence	.17	-.16
R^2_{adj}		.16	.08
Disgust	Kills	-.36*	-.23
	Deaths	.17	.24

	Violence	.15	.17
R^2_{adj}		.17*	.20*
Contempt	Kills	-.39*	-.04
	Deaths	.17	.18
	Violence	-.11	-7
R^2_{adj}		.14	.05
Fear	Kills	-.36*	-.19
	Deaths	.23	.11
	Violence	.10	-.02
R^2_{adj}		.15	.07

Note. * p<.05, ** p<.01, *** p<.001

Figure 1. Examples of the low violence versus high violence conditions.

Psychological Assessment

Press CRTT to Measure Aggressive Behavior: The Unstandardized Use of the Competitive Reaction Time Task in Aggression Research

Malte Elson, M. Rohangis Mohseni, Johannes Breuer, Michael Scharkow, and Thorsten Quandt

Online First Publication, January 20, 2014. doi: 10.1037/a0035569

CITATION

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014, January 20). Press CRTT to Measure Aggressive Behavior: The Unstandardized Use of the Competitive Reaction Time Task in Aggression Research. *Psychological Assessment*. Advance online publication. doi: 10.1037/a0035569

Psychological Assessment

1040-3590/14/$12.00 DOI: 10.1037/a0035569

Press CRTT to Measure Aggressive Behavior: The Unstandardized Use of the Competitive Reaction Time Task in Aggression Research

Malte Elson
University of Münster

M. Rohangis Mohseni
University of Osnabrück

Johannes Breuer
University of Münster

Michael Scharkow
University of Hohenheim

Thorsten Quandt
University of Münster

The Competitive Reaction Time Task (CRTT) is the measure of aggressive behavior most commonly used in laboratory research. However, the test has been criticized for issues in standardization because there are many different test procedures and at least 13 variants to calculate a score for aggressive behavior. We compared the different published analyses of the CRTT using data from 3 different studies to scrutinize whether it would yield the same results. The comparisons revealed large differences in significance levels and effect sizes between analysis procedures, suggesting that the unstandardized use and analysis of the CRTT have substantial impacts on the results obtained, as well as their interpretations. Based on the outcome of our comparisons, we provide suggestions on how to address some of the issues associated with the CRTT, as well as a guideline for researchers studying aggressive behavior in the laboratory.

Keywords: aggression, standardization, measurement, methodology, Competitive Reaction Time Task

Supplemental materials: http://dx.doi.org/10.1037/a0035569.supp

Researchers have been looking into predictors and antecedents of aggression for a long time, but the field is still struggling to find valid methods to measure aggressive behaviors in the laboratory. The problem of measurement already begins with finding an appropriate definition of aggression. Baron and Richardson (1994) defined aggression as any behavior that is intended to cause harm to another person who intends to avoid this harm. To distinguish between aggression and violence, Ferguson and Rueda (2009) pointed out that "violent behaviors . . . are typically restricted to acts which are intended to cause serious physical harm" (p. 121), while "aggression as a class of behavior is much broader than violent behavior and can include numerous acts . . . which are neither physically injurious nor illegal" (pp. 121–122). Similar to the difficulty of defining aggression, finding a valid, reliable, and ethically acceptable measure of aggressive behavior for laboratory research is an intricate undertaking.

Despite such difficulties, numerous studies have been conducted and published in which laboratory measurements are employed to provide evidence that a large variety of stimuli increase aggressiveness. These stimuli include displayed violence in digital games (Anderson & Dill, 2000) and gory passages from the Bible (Bushman, Ridge, Das, Busath, & Key, 2007) but also alcohol (Bond & Lader, 1986) and other psychoactive drugs (Bond & Lader, 1988). The test most commonly used in laboratory research on aggressive behavior in adults, especially in studies on media violence, is a modification of Taylor's (1967) Competitive Reaction Time Task (CRTT). In this article, we briefly describe this test and how it has been used and modified for aggression research. Afterward, by applying the different analyses to data from three independent studies that employed the CRTT as a measure of aggressive behavior, we show that the many analysis procedures that have been used for the CRTT can lead to contradictory results even within the same data set.

The Competitive Reaction Time Task

In the original version of the CRTT by Taylor (1967), the Taylor Aggression Paradigm, participants were led to believe that they would be playing a reaction time game against another participant

Malte Elson, Department of Communication, University of Münster, Münster, Germany; M. Rohangis Mohseni, Center for Information Management and Virtual Teaching, University of Osnabrück, Osnabrück, Germany; Johannes Breuer, Department of Communication, University of Münster, Münster, Germany; Michael Scharkow, Department of Communication Studies, University of Hohenheim, Stuttgart, Germany; Thorsten Quandt, Department of Communication, University of Münster, Münster, Germany.

The research leading to the results of Studies 2 and 3 has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant 240864 (SOFOGA).

Correspondence concerning this article should be addressed to Malte Elson, Department of Communication, University of Münster, Bispinghof 9-14, 48143 Münster, Germany. E-mail: malte.elson@uni-muenster.de

in which the winner of a round would punish the loser with an electric shock. Participants who lost a round would receive shocks of varying intensity and when participants won a round they could adjust the shock levels for their alleged opponents. The intensity level of the shock was used as the measure for aggressiveness. Recent adaptations of the CRTT allow participants to set the intensity (usually volume and/or duration) of a noise blast instead of an electric shock (e.g., Bushman, 1995), as they are easier to use and bring up fewer ethical issues (Ferguson & Rueda, 2009). As there is no real opponent, the sequence of wins and losses, as well as the settings "chosen" by the opponent, are typically randomized and preset. Generally, louder and longer noise blasts are considered indicators of higher levels of aggressiveness.

There has been disagreement among aggression researchers about the validity of the CRTT. This debate spans different issues, including alternative motives for high or low settings (e.g., reciprocity or deterrence), demand characteristics of the experimental situation, confusion of competition and aggression, construct validity and external validity, the (psychological) distance between participant and opponent, a lack of alternative (i.e., nonviolent) responses, and the absence of social sanctions for aggression. However, since the present study focuses on issues of standardization, we refer the reader to the available literature that discusses all of the topics outlined above (Ferguson & Rueda, 2009; Giancola & Chermack, 1998; Giancola & Zeichner, 1995a; Ritter & Eslea, 2005; Savage, 2004; Tedeschi & Quigley, 1996, 2000).

Clinical Relevance

Although we are aware that the CRTT has not been developed as a clinical instrument, the test and its psychometric properties are relevant to clinical research for two main reasons:

1. The CRTT (in one form or another) is being used to answer research questions in the area of clinical psychology. This includes investigations of social and cerebral response in criminal psychopaths (Veit et al., 2010); effectiveness of prescription drugs in reducing hostility in panic disorders (Bond, Curran, Bruce, O'Sullivan, & Shine, 1995); and the facilitation of aggression through various substances, such as alcohol (Bond & Lader, 1986; Pihl et al., 1995) or prescription drugs (Bond & Lader, 1988; Weisman, Berman, & Taylor, 1998). In some cases, practical recommendations for clinicians regarding the diagnosis (McCloskey, Berman, Noblett, & Coccaro, 2006) and treatment (Ben-Porath & Taylor, 2002) of patients are made based on results obtained with the CRTT. Given the impact of clinical research on the definition, assessment, diagnosis, and treatment of disorders in clinical practice, the importance of using objective, reliable, and valid measures cannot be overstated. Accordingly, Ferguson and Rueda (2009) warned against using the CRTT for clinical research and making any statements about clinical implications based on findings obtained with the CRTT (particularly versions that use noise blasts instead of electroshocks).

2. Results from nonclinical fields of study in which the CRTT is particularly popular, such as media effects research, are frequently generalized into the clinical realm. Bushman and Gibson (2011), for example, describe the CRTT as "a weapon that could be used [by the participants] to blast their partner" (p. 30). Bushman et al. (2007) found support for "theories proposed by scholars of religious terrorism who hypothesize that exposure to violent scriptures may induce extremists to engage in aggressive actions" (p. 204) in their study on university students' noise blasts after reading a passage from the Bible in which God sanctions violence. They conclude that "to the extent that religious extremists engage in prolonged, selective reading of the scriptures, focusing on violent retribution toward unbelievers instead of the overall message of acceptance and understanding, one might expect to see increased brutality" (p. 205). Some scholars also warn about a public health risk and equate the effects of media violence on aggression to those of substance use, abusive parents, or poverty (Anderson et al., 2010). Anderson, Gentile, and Buckley (2007), for example, consider violent video games as one of several risk factors that may cause aggressive, violent behavior and, in highly extreme and rare cases, even school shootings. Sometimes scholars make direct comparisons with other research areas and, e.g., claim that the effects of media violence are similar to those of smoking on lung cancer (Bushman & Anderson, 2001). Given that such statements are being made in the scientific literature, it comes as little surprise that similar conclusions have been found in the public debate. In a recent court case, a prominent media violence researcher was paid by the defense to serve as an expert witness and he argued that a teenage victim of a brutal multiple murder was the perpetrator of that murder, in part because he played digital games (Rushton, 2013). As the CRTT is used in a majority of experiments in this field (Anderson et al., 2010), and taking into account the lack of standardization and data supporting the validity of CRTT score interpretations (Ferguson & Rueda, 2009), such conclusions or comparisons are particularly inappropriate.

Standardization

Our primary concern is the lack of standardization of both test procedure and data analysis for the CRTT. The test has been used in many different versions, sometimes even by the same authors (Ferguson, 2011). Due to this lack of standardization, Ferguson, Smith, Miller-Stratton, Fritz, and Heinrich (2008) suggested a standard procedure for the CRTT. This version consists of 25 trials, and the sequence of losses and wins, as well as the computer's settings of volume and duration, is preset and randomized, so that the same pattern exists for every participant. The settings of the computerized opponent are always displayed to the participant after each trial. The average volume and duration settings are treated as separate and equal scores for aggressive behavior, although the authors express their concerns about using duration settings at all in future studies. Until now, however, only a few researchers have adopted the version proposed by Ferguson et al. (2008).

Inconsistencies are found in the procedure of the CRTT, as well as in the ways in which the CRTT data are analyzed by different (and sometimes even the same) authors. While the procedural aspects refer to the setup of the test, i.e., how the raw data are generated, the statistical differences refer to how the data are analyzed. Of course, the procedural decisions also affect the options for statistical analyses. The following section is meant to list the variations of both procedures and statistical analyses of the CRTT that can be found throughout the literature.

Bushman and Baumeister (1998) used the sum of volume and duration settings in the first of 25 trials. They provided the reason for choosing this procedure in a footnote stating that the first trial is the only noise setting that is "unprovoked," (p. 222) and that in all subsequent trials the participants' settings in the study converged with those of their opponents (i.e., there was no difference between conditions).

Anderson and Dill (2000) calculated separate aggression scores from average volume and duration settings for each possible trial outcome (winning and losing), resulting in a total of four scores for aggressiveness. They expected retaliatory behavior to be higher after losing a trial. In addition, they log-transformed the duration data because they were positively skewed, but reported and interpreted the original mean differences as being significantly different between conditions. Although three out of four parameters yielded nonsignificant differences, they asserted a "convergence of findings" (p. 787) in their discussion.

The use of four separate parameters is also different from the procedure of Lindsay and Anderson (2000), who multiplied volume with log-transformed duration settings. The average over 25 trials of those products was their measure for overall aggression. Carnagey and Anderson (2005) averaged the products of volume and the square root of duration to form a single "aggressive energy score" (p. 887). No reason is given for this other than the claim that this single score supposedly is a valid measure and that duration should be square rooted. By contrast, Anderson and Carnagey (2009) only allowed their participants to set the volume, and they used two indicators for aggression (mean volume and the total number of volume settings of 8 to 10 on a total scale of 1 to 10). The reasons given for using this number of high volume settings were that they were more clearly aggressive, more likely to instigate retaliation, and easier to communicate to nonexperts.

The same inconsistencies in CRTT procedure can also be observed in other publications. Bartholow, Sestir, and Davis (2005) multiplied the average volume and duration settings to form a composite aggressive behavior score. Although Bartholow, Bushman, and Sestir (2006) also used volume and duration settings, they standardized and summed the two parameters instead of multiplying them. Conversely, Sestir and Bartholow (2010) only analyzed average volume, not allowing their participants to set duration at all. Finally, Engelhardt, Bartholow, and Saults (2011) used a one-trial version of the CRTT because they wanted to eliminate reciprocal behavior and concerns over retribution. Unfortunately, they used the term "noise intensity" (p. 541) throughout their article without specifying whether this includes volume, duration, or both, and how this index was calculated.

In a two-phase version of the CRTT used by Bartholow and Anderson (2002), participants played two complete rounds with 25 trials each. During the first phase, only the opponent could set the duration and intensity of the noise blasts. During the second phase, roles were reversed so that participants could retaliate for the punishment they received in the first phase. The authors only considered average volume and the number of high volume settings because they did not find a significant effect on any of the duration measures. The reason they gave is that their participants supposedly ignored the duration option. Anderson et al. (2004) separated the volume settings of Phase 2 into four different scores: first trial and means of Trials 2–9, 10–17, and 18–25. They claimed this was a common procedure with the CRTT because early trials were supposedly more important since they were more likely to be used for retaliation. We do not wish to inappropriately speculate or judge why there was not always a rationale for modifications to the CRTT as there might be numerous reasons (e.g., spatial limitations in academic journals), but we believe that the lack of justifications for such alterations is a fact that warrants mentioning in a methodological critique of the test in question.

Sometimes the option of setting the volume and/or duration to zero as a way to act nonaggressively is provided. However, including zero values in mean calculations could potentially skew some of the aggression scores mentioned above. Including settings of zero as an option also raises further questions, for example, how to handle trials in which participants set only one of the two intensity parameters to zero.

From a methodological point of view, inconsistent procedures and analyses are highly problematic because they infringe upon the objectivity criterion of psychological diagnostic test theory. Simmons, Nelson, and Simonsohn (2011) pointed out that flexibility in data collection, analysis, and reporting in psychological research dramatically increases actual rates of false-positive findings. Moreover, if there is no standardized procedure for a test and no standardized way to process the raw data into a meaningful value, the question remains whether the unstandardized score really approximates the true value of the construct. Aggression scores that are calculated with different procedural versions of the same test become very difficult to compare. Under the assumption that all these different procedures and analyses are equally capable of measuring the construct of aggressiveness, it is unclear why so many versions exist.

We believe that most of the time modifications to the CRTT have been made in good faith and for valid reasons, for example, to extend previous findings or to answer novel research questions (such as which opponent noise pattern elicits more aggressive responses). Without a doubt, theory-driven modifications of a method such as the CRTT, with the aim of answering specific research questions, can contribute to the understanding of psychological processes and extend the area in which a certain test can be applied. However, many authors do not explain in detail why they decided on a specific test procedure or on the aggression score they calculated from the raw data. In many cases, it is not clear why a particular score should be more suitable than others to address the respective research questions. Sometimes, the decision to focus on one of many possible scores seems to have been made post hoc, not prior to data collection. Our concern is not that changes to the structural aspects of the CRTT will render it useless but that it might tap into different constructs and that the results will have to be interpreted accordingly.

Reactivity

Participants can hear (and see) their opponents' settings and are likely to react to this in their own choice of volume and/or duration. While most researchers use preset randomized settings, they do not use the same pattern in all studies. Without a standardized noise pattern, participants of different studies will most likely receive different noise blasts by their opponent, which might cause variations in their settings.

Even with a standardized preset pattern of wins/losses and noise intensities, it is likely that the alleged opponent's responses have an effect on the behavior and/or motives (e.g., retaliation) of the participants. This can lead to a lack of item homogeneity, as it is possible that not all trials actually measure the same variable (i.e., the same behavior). Researchers using the CRTT have approached this problem with different rationales, e.g., by focusing on the first trial, as it is the only nonreciprocal data point in the test's one-phase version (Bushman & Baumeister, 1998) or because it is the first chance to retaliate in the two-phase version (Anderson et al., 2004). In the latter study, the authors used two different noise patterns (ambiguous and increasing) to show that, in fact, there is an effect on the participants' responses. However, the impact of trial outcome and intensity patterns is usually not controlled for, and to our knowledge, no one has ever systematically checked how much intra- and interindividual variance can be explained just by the opponent's volume and/or duration settings.

Research Questions

Until now, there has been no study that addresses the aforementioned objectivity issues by systematically comparing the different analysis procedures for the CRTT. To address this gap in the literature, we used data from three separate studies that were conducted independently of each other by the authors of this article, and all used the CRTT as their measure of aggression.

The three experiments were selected because they all studied digital games as a potential cause of aggressive behavior. Media violence is a research field in which the CRTT is particularly popular (Anderson et al., 2010). There is a fairly large body of research indicating that this stimulus does cause differentials in CRTT scores. In fact, all but two of the studies that we cited in the section on standardization investigated effects of violence in digital games on postplay aggression. In addition, the authors of this article were involved in these studies and thus had full access to the raw data. Each study was conducted separately and two different research groups were involved in the studies: The second author of this article was in charge of data collection and the original analysis of Study 1 (Mohseni, 2013), the first author did the same for Study 2 (Elson, Breuer, Van Looy, Kneer, & Quandt, 2013), and the third and fourth authors were responsible for Study 3 (Breuer, Scharkow, & Quandt, in press). Originally, the studies were conducted to investigate the effects of media violence and not to evaluate the CRTT.

The results presented in this article come from a secondary analysis of the data of these three studies. We want to stress that this article is not concerned with the original hypotheses and results of each study, and it was not our aim to replicate effects across studies. In this article, we are interested in the issues of standardization and reactivity associated with the CRTT. Hence, we do not want to compare effects *across studies* but investigate whether different analysis procedures lead to different results *within each study*.

More specifically, we wanted to answer the following research questions:

RQ1. Do measures of volume and duration converge?

RQ2. Do different methods to calculate CRTT aggression scores yield comparable results?

RQ3. Do the settings of the simulated opponent explain parts of the variance in the settings of the participants?

Study Descriptions

In this section, we briefly describe the methods, materials, samples, and procedures for each of the three studies, as well as their main findings. The following section will only describe the methods of each study with a special focus on the version of the CRTT that was used. Readers interested in the results of the original studies should refer to the online supplemental materials or the individual publications (Breuer et al., in press; Elson et al., 2013; Mohseni, 2013).

Study 1

This study (Mohseni, 2013) addressed the question whether situations of violent helping behavior in digital role-playing games increase aggressiveness and helpfulness outside the game.

Stimulus materials and measures. To answer this question, two independent variables, namely, in-game violence and in-game help, were created by modifying the role-playing game *The Elder Scrolls IV: Oblivion* (Bethesda Game Studios, 2006). Participants had to solve a quest in the game either by using violence or by stealthily traversing the map (violent vs. nonviolent). The quest was either framed as helping behavior (quest giver had to be saved) or as a treasure hunt. Participants were assigned to one of four conditions: "rescue" (helping with violence), "kill" (violence only), "help" (helping only), and "treasure hunt" (no helping and no violence).

Aggressive behavior was measured with Bushman and Baumeister's (1998) version of the CRTT: The Competitive Reaction Task Reward & Punish (Version 3.2.5). Participants were able to set the volume and duration. Volume was calibrated using Bruel & Kjaers Sound Pressure Level Meter Type 2232, resulting in settings from 1 (55 dB) to 10 (95 dB). Participants could also choose a setting of 0 (0 dB). A win/loss pattern (12/13 trials) and the opponent's settings (increasing noise intensities over time) were predefined for all participants. Since Giancola and Zeichner (1995b) found that men are more likely to express aggression in intensity, while women are more likely to use duration, the measure for aggressive behavior in this study was the volume setting in the first round for male participants, and the duration setting in the first round for females.[1] The Aggressive Motives Scale (AMS; Anderson & Murphy, 2003; Anderson et al., 2004) is supposed to measure instrumental aggression (two items) and revengeful aggression (four items) during the CRTT and was used to assess the participants' motives. Helping behavior was assessed as the willingness to assist in another experiment (Greitemeyer & Osswald, 2010).

Procedure and sample. Before entering the lab, participants met a male confederate who was supposedly waiting for the experiment to start. The experimenter led the confederate to his cabin first and told him to wait for a second experimenter. After that, the participant was brought to a cabin and asked to sign an informed consent form. Participants were then taught how to play the CRTT, as recommended by Anderson et al. (2007, p. 65). Directly after that, participants started playing one version of *Oblivion* (see above). When the quest was solved, the experi-

[1] It should be noted, however, that Giancola and Zeichner found this sex difference for electroshocks and not noise blasts. Whether these two stimuli can be equated as aggressive measures remains subject to further research.

menter returned and told the participant that his opponent was ready so that the CRTT may now begin. After the CRTT, participants filled out several questionnaires. At the end of the experiment, participants were thanked, debriefed, and those who did not participate for course credit received a monetary compensation of €10 (about $13).

Participants (N = 216; 188 males and 28 females)[2] were German bachelor and master students from different faculties. The largest proportion of participants majored in cognitive science (12.6%), psychology (7.5%), and law (5.6%). Mean age in the sample was M = 23.48 years (SD = 3.21). Due to language problems and a technical issue, data from two male participants had to be discarded from further analyses, leaving a total of 214 participants distributed over all four game conditions (53 in the "rescue" and "kill" conditions, 54 in the "help" and "treasure hunt" conditions).

Study 2

While Study 1 was more concerned with violence and helping as a means to reach a goal, in the second experiment (Elson et al., 2013) displays of violence and the game speed of a first-person shooter were manipulated to assess their effects on aggressive behavior.

Stimulus material and measures. Two features of the game *Unreal Tournament 3* (Epic Games, 2007) were modified: displays of violence and game speed. Game speed was either set to the default value of 100% (normal speed) or to 140% (high speed). In the violent conditions, players wielded a grenade launcher and shooting opponents led to a gory death animation. In the nonviolent conditions, participants used a tennis-ball shooting Nerf gun. Aggressive behavior was measured with the standardized version of the CRTT with volume and duration settings (Ferguson et al., 2008) using the experiment software *Presentation* (Neurobehavioral Systems, 2010). However, instead of using dB values, the range of volume was determined in a prestudy in which 15 participants were asked to select a volume level on the Windows sound settings that was "almost unbearable." Taking that value as the maximum, the volume was scaled down linearly, resulting in settings from 1 to 10. The win/loss pattern (12/13 trials) and opponent settings were randomized for each participant. The mean of Volume × Duration over all trials was used as the measure for aggressive behavior.

Additionally, physiological arousal was measured by four different parameters. Two measures assessed autonomic responses: heart rate (HR) and skin conductance level (SCL). Two others were behavioral indicators: body movement and pressure exerted on mouse and keyboard.

Procedure and sample. Participants were assigned to one of four conditions (nonviolent vs. violent, normal- vs. high-speed). After entering the lab and signing the informed consent form, participants received a short briefing about the game's controls and objectives. After they finished three warm-up rounds, the experimenter started the main playing session that lasted 12 min. Afterward, participants were told that the second part of the experiment was about to begin, in which they would play 25 rounds of a reaction time game against a participant in another laboratory. Instructions were also presented on the computer screen before the first trial. After completing the CRTT, participants were thanked and debriefed. As an incentive, 40 computer games were raffled among the participants after completion of data collection.

Participants (N = 87; 60 males and 27 females) were undergraduate (79.4%) and graduate (4.6%) students from two large German universities and high school graduates (12.6%) who were about to enroll as university students. All participants were recruited via the online recruitment tool *Cortex* (Elson & Bente, 2009) and 3.4% did not indicate their current occupation. University students were recruited via mailing lists and came from different fields, the largest proportion being communication (20.7%) and psychology (13.8%). Mean age of the participants was M = 26.07 years (SD = 5.87). Due to technical difficulties, data from three participants had to be discarded from further analyses, leaving a total of 84, evenly distributed over all four game conditions (21 each).

Study 3

The third study (Breuer et al., in press) was originally conducted to test the frustration-aggression hypothesis (Berkowitz, 1989) in the domain of video games. Specifically, this study investigated effects of winning/losing and opponent behavior in a co-located multiplayer sports video game on negative emotions and aggression.

Stimulus material and measures. Participants played a match of the soccer video game *2010 FIFA World Cup South Africa* (EA Canada, 2010) against a male confederate with substantial practice. The confederate was instructed to either win or lose against the participant and to either be friendly and helpful or to (mildly) trash-talk while playing. The trash-talking was not scripted, as it was meant to be adaptive. The confederate was provided with a list of sample phrases that he was free to combine and adapt according to the course of the game. Sample statements included ironical remarks such as "nice pass," or snarky comments such as "that was easy."

The version of the CRTT used in this study differed from those used in Studies 1 and 2 in several aspects. To provide the participants with an appropriate target for their aggression, participants in Study 3 were told that they would play the CRTT against the same person against whom they had played the soccer console game. Instead of an auditory cue, participants had to react to a visual cue displayed on-screen in order to address the issue of the noise blasts being a potential means to reduce the reaction time of the opponent (Adachi & Willoughby, 2011). If they won a trial, participants could choose the duration of blasts, ranging from 1–9 s, using the corresponding number keys on a keyboard. There were a total of 10 trials and the sequence of winning and losing trials (five each), as well as the opponent's settings, were randomized individually for each participant. The volume setting was excluded because Ferguson and Rueda (2009) reported insufficient correlations between duration and intensity in their CRTT validation study, and duration (in seconds) was believed to be a more intuitively comprehensible unit for all participants. A very unpleasant noise level was determined in a small prestudy with five participants. This volume level was held constant in all trials. The

[2] In the original study, data of the female subsample were not used for analysis. Therefore, results presented here may differ from results of the original study.

duration setting of the first trial was used as the measure for aggressive behavior. Afterward, the CRTT participants also completed a brief questionnaire that included items on negative emotions. Negative emotions were measured by four items from the German translation (Krohne, Egloff, Kohlmann, & Tausch, 1996) of the Positive Affect Negative Affect Schedule (Watson, Clark, & Tellegen, 1988).

Procedure and sample. Upon arrival at the laboratory, each participant gave their informed consent and was asked whether he or she had any experience playing *2010 FIFA World Cup South Africa* on the Xbox 360 console. If the participant did not have any experience with the game, he or she played a practice session against an easy computer opponent for 5 min.

Following the practice phase, participants played a 2 × 5 min match against the confederate in one of the four conditions. After the match, the confederate was led into an adjacent room, and participants were told that they would play a reaction time game against him. At the end of the experiment, participants were thanked and those who did not participate for course credit received a monetary compensation of €10 (about $13). All participants were debriefed via e-mail at the end of the data collection phase to avoid an early uncovering of the role of the confederate and the purpose of the study.

A total of 91 participants signed up for the experiment using the *Cortex* online recruiting tool (Elson & Bente, 2009). The majority of the sample (80.3%) was bachelor and master students, and most of them were enrolled for communication (49.9%). The data of 15 participants were excluded from further analysis because of language problems, participants having suspicions about the purpose of the study, knowing the confederate, or the game resulting in a draw. The mean age of the remaining 76 participants (48 female; 28 male) was $M = 22.60$ years ($SD = 3.20$).

Secondary Analysis of the CRTT Data

In order to answer the first research question (*RQ1*), asking if volume and duration measures converge, we calculated the correlation between these two variables for Studies 1 and 2. In Study 1, volume and duration measures for each trial showed a medium-sized significant correlation ($r = .49$), p (one-tailed) $< .001$. The correlation of average volume and duration measures for each participant was substantially higher ($r = .82$), p (one-tailed) $<$ .001. Study 2 showed a similar medium-sized significant correlation for each trial ($r = .46$), p (one-tailed) $< .001$, while the correlation of average volume and duration measures for each participant was higher ($r = .76$), p (one-tailed) $< .001$.

Convergence of CRTT Scores

Although the correlations per participant were quite high and show that volume and duration measures are related, they might not measure the same construct. The results of our comparisons between different analysis procedures support this notion, as we show in the remainder of this section. In our comparisons, we included most of the analysis procedures for the one-phase CRTT that have been published previously and could be applied to our data sets.

Naturally, it would have been possible to add even more aggression scores to our analyses or to use them in more advanced statistical analyses, such as multilevel models or latent growth models. However, we wanted to test whether the available scoring techniques are comparable and whether they lead to different results in the most basic and, in this case, the most common statistical analyses. Adding yet another method for scoring would only further complicate any attempt at standardization. For our comparisons, we only used one independent variable per study. For Studies 1 and 2, this is violent content; for Study 3, we focused on the outcome of the game (winning vs. losing). These variables were selected because they have been previously identified as causes of aggressive behavior (as reflected by CRTT scores). As stated before, the aim of this analysis is not to examine the convergence of the effects of the independent variables on the CRTT *across studies*. Instead, we are interested in the variability of results when using different CRTT scores *within each study* and whether this variability can be replicated in studies that differ in their design.

Study 1. For Study 1, none of the aggression scores calculated were significantly different between the experimental conditions (see Table 1). In-game violence only had a small, but nonsignificant effect on the volume chosen in the first trial. However, it should be noted that there was a high variability in the indices' significance levels (from $p = .070$ to .934).

Study 2. The comparisons for Study 2 show a more ambiguous pattern (see Table 2). With the exception of average volume after losing, all measures for aggression based on average intensity (volume and duration) yielded nonsignificant differences between the experimental conditions with p-values ranging from .045 to .668. The range of p-values between all indices (including those not based on averages) was between $<.001$ and .959. The effect on volume was mostly larger than on duration, indicating that volume and duration are at least not equal in terms of their sensitivity to the effects of stimuli. This difference could not be observed for the first trial settings, however. Effect sizes ranged from .0 to .39, the largest being the number of high volume blasts, $F(1, 80) = 16.01$, $p < .001$, $\omega = .39$, showing that participants who played a violent game used volume settings from 8 to 10 more frequently than those that played a nonviolent game. This suggests that the number of extreme volume scores does not tap into the same construct as mean-based scores.

Table 1
Study 1: Effects of In-Game Violence on Different CRTT Aggression Scores

Aggression score	$F(1, 210)$	p	ω
Mean volume	1.23	.269	.03
Mean volume after wins	1.10	.295	.02
Mean volume after losses	1.06	.303	.02
Mean duration	0.23	.633	.0
Mean duration after wins	0.01	.934	.0
Mean duration after losses	0.59	.443	.0
Mean volume × duration	0.63	.429	.0
Mean volume $\times \sqrt{\text{duration}}$	0.83	.364	.0
Mean volume × $\log_e$(duration)	2.31	.130	.08
Total high volume settings	0.13	.719	.0
Total high duration settings	0.1	.730	.0
First trial volume	3.31	.070	.10
First trial duration	0.26	.609	.0

Note. CRTT = Competitive Reaction Time Task.

Table 2
Study 2: Effects of Displayed Violence on Different CRTT Aggression Scores

Aggression score	F(1, 80)	p	ω
Mean volume	3.28	.074	.16
Mean volume after wins	1.46	.230	.07
Mean volume after losses	4.14	.045	.19
Mean duration	0.95	.334	.0
Mean duration after wins	0.19	.668	.0
Mean duration after losses	1.74	.191	.09
Mean volume × duration	2.78	.099	.14
Mean volume × $\sqrt{\text{duration}}$	2.77	.100	.14
Mean volume × $\log_e$(duration)	2.17	.144	.12
Sum high volume settings	16.01	<.001	.39
Sum high duration settings	0.17	.685	.0
First trial volume	0.08	.779	.0
First trial duration	0.01	.959	.0

Note. CRTT = Competitive Reaction Time Task.

Mathematically, it seems odd that the significantly larger number of high volume settings did not cause the mean volume to be significantly higher. To explain this puzzling finding, we tested whether there were also differences in the number of low volume settings (range 1–3). A two-factorial analysis of variance (ANOVA) revealed a significant effect of violent game playing on the total of low volume settings, $F(1, 80) = 10.57$, $p = .002$, $\omega = .32$, indicating that playing a violent game also led to the more frequent use of volume settings from 1 to 3. Apparently, high volume and low volume settings canceled themselves out, resulting in nonsignificant mean differences.

Study 3. Since participants were only able to set duration in Study 3, we could use only six out of 13 analysis procedures (see Table 3). While there were small effects of losing on mean duration and mean duration after wins, they were both nonsignificant. The only significant effect of game outcome was on duration settings in the first trial ($\omega = .20$, $p = .043$). Participants who lost against the confederate on average chose longer durations. The remaining aggression scores yielded small effects ranging from $\omega = .09$ to .15, with p-values between .096 and .212.

Convergence of the different scores. Looking at the comparisons of analysis procedures for each study, the answer to the second research question (*RQ2*), asking whether different methods to calculate CRTT scores produce comparable results, has to be negative. Effect sizes ranged from $\omega = .0$ to .10 in Study 1, .0 to .39 in Study 2, and .09 to .20 in Study 3. As effect sizes are not the same for all analysis methods, it seems that the calculation of different aggression scores can lead to results that are substantially different from each other; in one case, even diametrically opposed. For example, depending on which aggression score is calculated (and reported) with the data from Study 2, results could provide evidence that playing a violent digital game increases postplay aggressive behavior (number of high volume settings), decreases it (number of low volume settings), or has no effect at all (most other indicators). We also found inconsistencies in Study 3, with one variant yielding a significant effect, while five others did not. Note that Study 1, having the largest *N* by far, yielded the most consistent effect sizes. This could indicate an issue with precision in the CRTT, meaning that the test is prone to producing false-positive significant differences when conducted with small to medium samples as is common practice in experimental psychology (see Simmons et al., 2011). However, even in Study 1 there was a large variability in p-values from .070 to .934.

While none of the scores in Study 1 reached the criterion of significance, using the methodological flexibility procedures discussed by Simmons et al. (2011), such as adding or removing a small number of participants or controlling for simple covariates (e.g., age), some of the findings could easily become significant. Not only are those practices relatively common in psychological research (John, Loewenstein, & Prelec, 2012), statistically significant findings also dramatically increase the likelihood of publication, resulting in a peculiar prevalence of p-values just below .05 (Masicampo & Lalande, 2012), particularly in controversial fields with major social implications, such as media violence (Ioannidis, 2005). Of course, we do not wish to imply that aggression research (or other psychological research) is invalid on these grounds, but want to express our concern that the unstandardized use of the CRTT might increase the likelihood of finding (and publishing) significant effects when, in fact, there are none.

Reactivity

To answer the third research question (*RQ3*), asking whether the settings of the simulated opponent can explain parts of the variance in the settings chosen by the participants, we regressed the participants' volume and duration settings on the results of the previous trial (win/loss, volume and duration of the noise blast received). As win/loss pattern and opponent's intensity settings were identical for all participants in Study 1, we analyzed the intraindividual variation of all respondents across 25 trials for this study by regressing the participants' settings in every trial on the settings of the opponent in the previous trial. On average, the explained variance in the respondents' settings was $R^2 = .39$ ($SD = .25$) for volume and $R^2 = .32$ ($SD = .21$) for duration. Effectively, this means that about one third of all variance in the CRTT data could be explained solely by the opponent's settings in the previous round. Compared to Studies 1 and 3, the amount of variance explained just by trial outcome and the opponent's settings of the previous round was a lot smaller in Study 2: $R^2 = .13$ ($SD = .09$) for participant's volume settings and $R^2 = .16$ ($SD = .12$) for duration settings. In Study 3, both outcome and noise duration were randomized in every trial individually for each participant. As the settings of the opponent were only displayed after loss trials, we could only estimate the amount of variance explained by the opponent's settings for these trials. The results were very

Table 3
Study 3: Effects of Game Outcome on Different CRTT Aggression Scores

Aggression score	F(1, 72)	p	ω
Mean duration	2.84	.096	.15
Mean $\log_e$(duration)	2.79	.099	.15
Mean duration after wins	2.84	.096	.15
Mean duration after losses	1.63	.205	.09
Sum high duration settings	1.58	.212	.09
First trial duration	4.24	.043	.20

Note. CRTT = Competitive Reaction Time Task.

similar to those of Study 1, with an average of $R^2 = .41$ ($SD = .32$) of explained variance in the participants' duration settings (after loss trials).

To further investigate whether there would be differences in the participants' reactivity depending on the outcome of a previous round, we conducted separate regression analyses for trials after rounds won versus rounds lost (see Table 4) for Studies 1 and 2. Study 3 was excluded from this comparison as the opponent settings were not displayed when a participant won a round. As in our previous analyses, there were large differences in the reactivity between Studies 1 and 2. More important, there was an influence of opponent settings independent of the outcome of the previous round, with slightly inconsistent R^2s ranging from .34 to .47 for Study 1, and .16 to .22 for Study 2.

Of course, the reactivity problem might even be more complex, as these analyses only accounted for the settings of the previous round on the next one. It is certainly possible that the opponent's settings also produced changes in reactivity in subsequent rounds.

Motives

The issue of reactivity also opens up additional questions about what influences the participants' settings in the CRTT. As mentioned above, other studies found differences in responses depending on the characteristics of the noise pattern. However, there have only been a few attempts to identify what causes those reactions, such as personality traits or emotional states. We believe that motivations to aggress also play a major role here, as these motivations could be influenced by violent media content, personality traits, and the behavior of the opponent.

Participants in the first study were asked about their motivations during the CRTT. We used these data to determine the amount of variance in the participants' settings explained by different motives. While the AMS is supposed to measure the motives of instrumental and revengeful aggression, the phrasing of the items is open to alternative interpretations. For instance, the first item of the instrumental aggression scale, "I wanted to impair my opponent's performance in order to win more," could measure a nonaggressive competition motive, while the second item, "I wanted to control my opponent's level of responses," could tap into the motive to control or manipulate the opponent. Within the revengeful aggression scale, the two items, "I wanted to pay back my opponent for the noise levels he or she set" and "I wanted to blast him or her harder than he or she blasted me," do seem to measure a motive for revenge, but the other two, "I wanted to make my opponent mad" and "I wanted to hurt my opponent," could measure a form of aggression that has nothing to do with revenge. Additionally, a scale intercorrelation of up to $r = .49$ (Anderson et al., 2004) may indicate that the AMS only measures one single construct.

As the factorial structure by Anderson and Murphy (2003) could not be replicated (since the two items of the instrumental aggression subscale correlated negatively), all items were subjected to a principal component analysis (PCA) with orthogonal rotation,[3] which revealed two different factors. The first consisted of three items (impair performance, make mad, hurt), and the second consisted of two (control response, pay back), leaving one item that loaded on both factors (blast harder). The first factor seems to reflect the motive of aggression and competition, while the second seems to reflect retaliation and control. Accordingly, these factors cannot be interpreted as individual motives. They are still useful for predicting aggression scores though, especially since using all items together in a single linear regression model would very likely lead to issues of multicollinearity.

Theoretically, the motives could both moderate and mediate the effect of any independent variable on the CRTT scores. As the wording of the AMS items clearly shows, some or even most of the motives have a trait component. There are people who are generally more likely to (re)act aggressively or to retaliate. The overlap with influential personality traits, such as trait aggression, would speak for the motives being moderators. At the same time, the motivations also have state components as they can be influenced by both the treatment (e.g., retaliation in the case of a rude opponent in Study 3) and the CRTT itself (settings chosen by the opponent; see section on reactivity). For example, aggressive motives should be more prominent in aggressive conditions and result in higher CRTT scores, while nonaggressive motives should be more prominent in nonaggressive conditions and result in lower CRTT scores.

Regressing the motives on "mean volume × duration" in Study 1, both factors led to higher mean settings, explaining 47% of the variance (see Table 5).

The first factor also predicted higher settings in the first trial, whereas the second was not significant here. In this case, both factors explained 26% of the variance (see Table 6). Although there was no interaction between the effects of experimental conditions and motives, this does not necessarily mean that the motives were not influenced by any of the independent variables. As our data show, these motives can have a large impact on CRTT scores. However, further research is needed to investigate systematically whether the AMS really measures motives (as they are assessed post hoc), whether other motives could play a role (e.g., conformity or social desirability), and in which way other variables (e.g., personality traits, or experimental stimuli) possibly influence these motives.

Table 4
Effects of Opponent's Settings on Participant's Settings by Outcome of the Previous Round

		R^2			
		Study 1		Study 2	
Participant settings	Previous round outcome	*M*	*SD*	*M*	*SD*
Volume	Win	.39	.25	.21	.17
	Lose	.47	.27	.16	.13
Duration	Win	.34	.22	.22	.18
	Lose	.39	.24	.18	.16

[3] Conducting a principal axis factoring instead of a PCA and/or using oblique instead of orthogonal rotation, led to the exact same factor structure (with some items loading lower on their corresponding factor in cases of the PAF, and with small interfactor correlations of $r < .12$ in cases of oblique rotation).

Discussion

Our results show that modifications to the CRTT test procedure result in different assessments of behavior. For example, whether zero is included as an option for participants to set their intensity and whether the pattern of opponent settings is increasing or ambiguous, might lead to very different findings, even if the studies are otherwise similar. While it might seem trivial that different versions of a test yield different results, this is a major problem in the case of the CRTT as it is being used widely, and as there is still no thoroughly validated standardized test procedure.

Methodologically even more problematic, however, are the differences among analyses of the raw data collected with the CRTT. The choice of a calculation method for aggression scores can severely influence the significance, size, or even direction of an effect. Our data could support different hypotheses about the effects of our stimuli, by either deliberately omitting some of the results, or just by picking one aggression score at random.

Furthermore, our findings suggest that volume and duration do not measure the same construct, although they clearly seem to be related. This does not necessarily constitute a problem with the CRTT. In fact, we would consider it a benefit if the CRTT captured different (sub) dimensions of aggressive behavior. However, no attempts to systematically disambiguate the different latent variables supposedly measured by volume and duration have been made thus far. Using the electroshock variant of the CRTT, Verona, Reed, Curtin, and Pole (2007) defined shock intensity as overt and shock duration as covert aggression and found differential use of the two settings between men and women (see also Giancola & Zeichner, 1995b; Ritter & Eslea, 2005). Whether these initial results also apply to the noise blast version of the CRTT, however, has yet to be tested. Thus, sex could be a further variable to be considered when using the CRTT.[4] The main concern of the present study is to compare different CRTT scores within the studies. The proportion of men and women in each study could explain the differences in effect sizes across studies, but not the variability within each one. The results presented in this article support the question raised by Ferguson et al. (2008), asking whether the CRTT should be used in aggression research at all. In order to find a definitive answer to this question, however, further studies evaluating the objectivity, reliability, and validity of the CRTT are necessary.

The practical implications of our results are that one has to consider and interpret findings of studies in which the CRTT is used with extreme caution, particularly when detailed reasons for any modification are not provided. We do not want to suggest that the consumption of drugs, alcohol, and violent media does not induce or increase aggressiveness. In fact, we believe that any research looking into causes of human aggression is of high relevance, as long as the results are reliable. The problems associated with the use of the CRTT, however, can seriously diminish the credibility and significance of any laboratory research on aggression.

Table 5
Study 1: Effects of Motives on Mean Volume × Duration

Factor	*B*	*SE B*	β	*p*
1	8.34	0.639	.65	<.001
2	2.66	0.639	.21	<.001

Note. $R^2 = .47$ ($p < .001$), $R^2_{\text{adjusted}} = .47$, variance inflation factor = 1.00.

Table 6
Study 1: Effects of Motives on First Trial Volume × Duration

Factor	*B*	*SE B*	β	*p*
1	5.26	0.616	.51	<.001
2	−0.56	0.616	−.05	.366

Note. $R^2 = .26$ ($p < .001$), $R^2_{\text{adjusted}} = .25$, variance inflation factor = 1.00.

Of course, our research has its own limitations that should be acknowledged. First, only data sets from three studies were available to us, while the number of experiments using the CRTT is much larger. Replicating our comparisons with more and possibly even larger data sets could help to identify a potential bias in the different ways used to process raw CRTT data.[5]

Another limitation is that all three studies used digital games as stimuli to elicit aggression. The consolidation of our findings would require supplemental analyses of data sets from studies that investigate other possible sources of aggression. Ironically, the three studies presented in this article all employed different versions of the CRTT. For Study 1, Version 3.25 of the Competitive Reaction Task Reward & Punish by Bushman and Baumeister (1998) was used. Study 2 used the standardized version suggested by Ferguson et al. (2008). Study 3 made modifications to Ferguson's version due to the nature of the research questions and hypotheses, although one might argue that the exclusion of volume was not directly justified by the research interest. The choice to only use duration was mainly made for practical reasons, such as the problem of proper calibration (see section on calibration and scaling below). All three studies also originally used different scores as the measure for aggressive behavior (first trial volume and duration, mean volume and duration). An ideal comparison among the studies would make sure that they all used the same (standardized) version of the CRTT. However, the differences in this set of studies also reflects the current practice of research using the CRTT and suggests that differences in CRTT procedure (e.g., noise pattern) might cause differences in the results.

Despite various problems with the CRTT and its analysis, there are currently no suitable alternatives for a behavioral measure of aggression that can be used in laboratory research with adult participants. Some scholars consider the Hot Sauce Paradigm (HSP; Lieberman, Solomon, Greenberg, & McGregor, 1999) to be a better measure (Adachi & Willoughby, 2011). However, a systematic validation of this test has only begun quite recently (e.g., Beier & Kutzner, 2012).

If we do not want to abandon laboratory research on aggressive behavior altogether, we have to find ways to deal with the problems of available tests. The following two sections are meant to

[4] A systematic examination of sex differences could not be made with the three data sets available to us, as the ratios across sex and the study designs differed too much between the studies.

[5] Authors willing to share their CRTT raw data with us for further analysis are welcome to contact the first author of this article.

provide some recommendations on how to use the CRTT in future studies and offer suggestions on how it might be improved upon or complemented by other methods. In the first section, we present recommendations for the future use of the CRTT based on the findings of the present study, while the second part is concerned with general advice from previous studies and our own considerations.

Recommendations Based on Our Findings

Based on our findings, we want to provide guidelines for researchers studying aggressive behavior in the laboratory. Following the structure of the article, we offer suggestions on how to address issues of standardization and how to control reactivity problems.

Standardization

At this point it is impossible to say which CRTT variant is the "right one." We recommend defaulting to the standardized version of the CRTT suggested by Ferguson et al. (2008), simply because it is relatively close to the original CRTT by Taylor (1967), without having the ethical problem of using electroshocks. Of course, without proper validation, any variant is technically as good as the next one. Nonetheless, we still think the version suggested by Ferguson et al. (2008) provides a good starting point, as its authors were the first who explicitly called for standardization and suggested a "standard operating procedure." If there are reasons for modifications, researchers should provide these and explain the benefits of their changes in detail. In keeping with the suggestions of Simmons et al. (2011), authors should ideally also explain if and how the results of their studies could differ using the standardized version of the CRTT.

With regard to the analysis, there is no definitive answer to the question of how to calculate aggression scores, or whether different scores might measure different types of aggression, as long as none of them have been properly validated. As it seems that volume and duration do not measure the exact same construct, it is advisable to consider them as separate measures for related subdimensions of aggression. If researchers are interested in measuring unprovoked aggression, they should also look at the settings in the first trial. Those studying provoked aggression or retaliation, on the other hand, should focus on all trials except the first one.

We are aware that modifications to established methods (e.g., the Stroop test) for new research questions are not uncommon in psychological science. However, we see two major differences compared to the use of the CRTT: First, the modifications being made are rarely theory-driven. Often, new variants seem to include arbitrary changes that are neither explained nor justified on the grounds of new research questions. Second, many of the variations of the CRTT are generalized to real-world acts of aggression, although there are no data supporting the external validity for any one of them. We believe that before a laboratory procedure is modified to answer particular research questions, there should be compelling evidence for sound psychometric properties of a "base version" of that procedure from which the variations can be derived. This is particularly true when researching a clinical topic that is highly relevant for society, such as aggression.

Reactivity

Should researchers decide to use all trials, they should analyze and report how much of the variance in the participants' settings can be explained by the settings of the opponent in the previous round. We do not necessarily consider reactivity a major problem, as it confirms that, in principle, the CRTT measures some kind of reciprocal behavior by the participant. One should not forget, however, that the opponent noise pattern is pregenerated and does not adapt to participants' settings, meaning that it is arguably not reciprocal and, hence, artificial and not very human-like. The reactivity might also become an issue when investigating the effects of antecedent stimuli, as they could potentially interact with reactivity effects, or even be superimposed by them. As this is the nature of the CRTT (with the exception of the first trial), we strongly recommend reporting the magnitude of the reactivity in a similar fashion as in this article.

To understand why participants chose their settings, the CRTT should be followed by an assessment of the participants' motivations, such as the AMS (Anderson & Murphy, 2003), although it remains unclear if this measurement of intentions is influenced by the CRTT itself. The AMS should be extended to assess more motives, such as conformity, deescalation, or schadenfreude, as these could also play a role. Studies on provoked/motivated aggression (e.g., the frustration-aggression hypothesis) should provide participants with an appropriate target for their aggression (e.g., another participant or confederate who provokes them). Using identical patterns of wins/losses and opponent settings for all participants is another important measure to address standardization and reactivity issues. We would even go one step further and endorse a standardized pattern for *all studies* using the CRTT (unless, of course, the pattern is being manipulated as an independent variable), as we consider this a part of the test material that should not introduce an uncontrolled source of variance.

General Recommendations

Nonaggressive Options

If zero settings as a nonaggressive option for participants are included, considering how to properly treat them in the statistical analyses is paramount. However, we remain skeptical whether zero settings can actually be included without skewing mean-based aggression scores in further analyses, and therefore recommend using settings from 1 to 10. Instead of zero settings, following the suggestions of Tedeschi and Felson (1994), we advocate including other nonaggressive alternatives, such as options to interact with the alleged opponent or to revoke any exposure to a potentially aggressive situation (i.e., the option to avoid or abort the CRTT).

Calibration and Scaling

Another general issue of the CRTT's standardization concerns is its calibration and scaling. Typically, noise intensity can be set on a scale from 1 to 10 in steps of 5 dB, ranging from 50 to 95 dB, but sometimes up to 115 dB. Some versions of the test also offer the additional option of setting the volume to zero (0 dB). The decibel scale, however, is a questionable choice for scaling noise intensity for two reasons: first, standardizing dB levels in all types of headphones is highly susceptible to interference due to variances in the sound card producing, the headphone speaker playing, and the volume meter recording the noise. Second, it appears that most scholars using the CRTT consider the available intensity settings from 0 to 10 to reflect linear intervals in the volume and its unpleasantness for humans. However, decibel is a logarithmic unit. For example, the increase in discomfort from 50 to 55 dB is substantially smaller than the one from 90 to 95 dB. Moreover, the "maximum unpleasantness" varies greatly between studies, from 95 dB (a subway train at a distance of 60 m) to 115 dB (a rock concert); the latter actually being four times louder than the former. Without any preceding transformation of the data into a linear scale, the assumption of equidistance between levels of noise and unpleasantness can lead to a misestimation of resulting aggression scores.

Ideally, volume should be calibrated using high quality level meters (e.g., Bruel & Kjaer Type 2232). For calibration, *sone* should be preferred over decibels, as it is a better approximation of perceived loudness. The typical range of 55–95 dB roughly equals 5–105 sone. If it is not possible to calibrate in sone, decibels can be transformed to approximate perceived loudness as follows: $x = 2^{\Delta L/10}$, where x is the difference in loudness and ΔL is the difference in volume. For example, compared to 50 dB, 55 dB ($\Delta L = 5$) is perceived to be 1.4 times, 60 dB ($\Delta L = 10$) 2.0 times, and 95 dB ($\Delta L = 45$) 22.6 times louder. Calibrating headphones is tricky, as there exists no standardized calibration distance. Normally, decibels are measured with a distance of 1 m (~3.3 feet) from the source of the sound. However, this is of little practical use in the case of headphones. We therefore suggest putting the volume meter as close to the headphone as the human ear would be.

Validity

While some authors found proof for the *convergent and discriminant* validity of the CRTT (Anderson & Bushman, 1997; Carnagey & Anderson, 2005; Giancola & Chermack, 1998), others did not (Ferguson, 2007; Ferguson & Rueda, 2009; Ferguson et al., 2008; Tedeschi & Quigley, 1996, 2000). This dissent is not so much based on different findings as it is on different convictions about which findings, indicators, or forms of validation are suitable to prove validity. For instance, some authors (e.g., Giancola & Chermack, 1998) believe that because aggressive persons score higher in the CRTT than nonaggressive persons, this would indicate the external validity of the test. According to Tedeschi and Quigley (2000), the reasoning behind this form of validation is that "If aggressive people produce behavior A more often than nonaggressive people outside the laboratory, and aggressive people produce behavior B more often than nonaggressive people inside the laboratory, then A = B" (p. 133). Specifically, if aggressive people behave violently more often than nonaggressive people in the "real world," and aggressive people give higher noise blasts than nonaggressive people when using the CRTT in the lab, then the noise blast intensity should be an indicator of violent behaviors. However, according to Tedeschi and Quigley, this conclusion constitutes a logical fallacy, which they illustrated with another example: High temperatures cause people to drink more fluids outside a laboratory and to rate others more negatively inside a laboratory. However, this does not allow the conclusion that drinking more fluids is the same kind of behavior as giving negative ratings.

In their extensive literature review of clinical and research aggression instruments, Suris et al. (2004) were unable to find a construct validation study for the CRTT. A recent lab–field comparison from 82 meta-analyses by Mitchell (2012) raises further concerns about the *external* validity of laboratory aggression research in comparison to other areas of psychological research, particularly industrial-organizational psychology.[6]

In line with Tedeschi and Felson (1994), we are convinced that an intent to harm (physically or otherwise) is a necessary and indispensable precondition for every form of aggressive behavior. Therefore, the main problem of all aggression measures is how to ensure that participants really have the intent to harm. Within the CRTT, plausible nonaggressive explanations for administering strong noise blasts could, e.g., be (a) falling for the cover story or (b) the desire to satisfy the experimenter and, thus, the demand characteristics of the test, especially if there is no nonaggressive option (Ritter & Eslea, 2005). Motive(s) determining the strategy in the CRTT procedure can also lead to very different reactions to the opponent's noise pattern. The underlying strategy or motive essentially determines the response pattern in laboratory measures of aggression. We believe it is paramount to gather evidence for the CRTT's *construct* validity in order to satisfy the definition of aggression and to ensure that participants actually have the intent to harm their alleged opponents. Asking participants about their intentions right after the CRTT with the AMS (Anderson & Murphy, 2003) has its limitations due to their retrospective nature and possible biases, such as social desirability. Therefore, validation studies in which participants' intentions are uncovered outside of an experimental situation should be given preference. This could be achieved by a study similar to Beier and Kutzner's (2012) validation of the Hot Sauce Paradigm (Lieberman et al., 1999) and with the inclusion of (additional) qualitative methods, such as interviews.

If the CRTT is used in experimental studies that are not (exclusively) designed to evaluate the validity of the test, researchers should include additional (ideally validated) measures of aggressiveness to investigate convergent and discriminant validity. To ensure that differences between experimental groups observed with the CRTT can be interpreted accurately, measurement of aggressive personality traits and pre-post designs can be viable solutions (Valadez & Ferguson, 2012), although pre-post designs could again introduce the problem of sensitization and contamination of the second measurement.

A further validation instrument could be physiological responses to the CRTT. In many studies that investigate causes of aggression, arousal indicators (most often galvanic skin response and heart rate) are frequently recorded during the stimulus exposure but not during the CRTT. Physiological responses could contribute to the test's validity, as more aggressive participants should show increased arousal (Zillmann, 1983). This would also help to uncover which events (losing a round, receiving or setting noise blasts) in the CRTT are particularly agitating, although

[6] Although it should be mentioned that there are domains (e.g., gender-focused comparisons) in which matters seem even worse.

naturally, physiological measurements do not allow drawing conclusions about an intention to hurt.

To summarize the order in which our recommendations should be considered, we created a flowchart (see Figure 1).

Conclusions

Our study provides empirical evidence that the CRTT suffers from several objectivity issues, in particular a lack of standardization. This can lead to inaccurate statements about the effects of various stimuli used in aggression research, such as violent media or alcohol. However, this article only addresses problems associated with the objectivity of the test. Despite the impact that this has on the significance of laboratory research on aggression, we believe that a satisfying solution to the problem of objectivity is feasible. Based on our own findings and previous methodological work on the CRTT, we can draw three main conclusions:

1. The results of studies that use the CRTT and meta-analyses that include these have to be interpreted with great caution. Moreover, given the questionable external validity of the test, researchers should be careful when they generalize results to situations outside the lab or make inferences about potential long-term effects to the point of public health issues.

2. Scholars who still want to use the CRTT to measure aggression must agree on a standardized version of the CRTT (as proposed by Ferguson et al., 2008). All modifications to this standardized version should be made explicitly and justified properly in the corresponding research reports.

3. If a standardized version is established and all issues of objectivity are addressed, the next step is to use this version for thorough validation studies that also investigate the construct validity of the test.

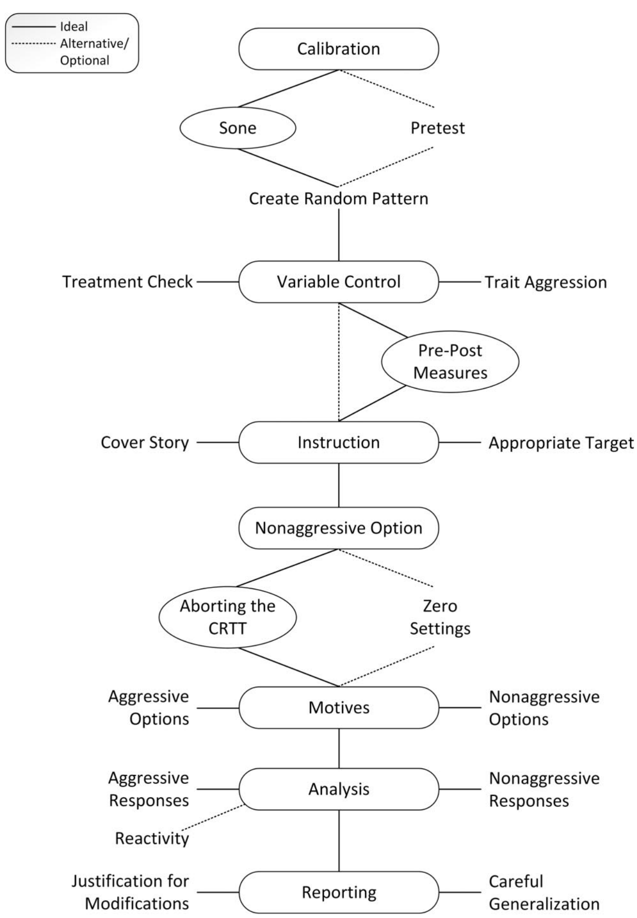


Figure 1. Recommendation flowchart for using the Competitive Reaction Time Task (CRTT).

References

Adachi, P. J. C., & Willoughby, T. (2011). The effect of video game competition and violence on aggressive behavior: Which characteristic has the greatest influence? *Psychology of Violence, 1,* 259–274. doi:10.1037/a0024908

Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1,* 19–41. doi:10.1037/1089-2680.1.1.19

Anderson, C. A., & Carnagey, N. L. (2009). Causal effects of violent sports video games on aggression: Is it competitiveness or violent content? *Journal of Experimental Social Psychology, 45,* 731–739. doi:10.1016/j.jesp.2009.04.019

Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin, A. J., Eubanks, J., & Valentine, J. C. (2004). Violent video games: Specific effects of violent content on aggressive thoughts and behavior. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 199–249). New York, NY: Elsevier. doi:10.1037/e552522012-008

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78,* 772–790. doi:10.1037/0022-3514.78.4.772

Anderson, C. A., Gentile, D. A., & Buckley, K. E. (2007). *Violent video game effects on children and adolescents: Theory, research, and public policy*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195309836.001.0001

Anderson, C. A., & Murphy, C. R. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior, 29,* 423–429. doi:10.1002/ab.10042

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., . . . Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin, 136,* 151–173. doi:10.1037/a0018251

Baron, R. A., & Richardson, D. R. (1994). *Human aggression* (2nd ed.). New York, NY: Plenum Press.

Bartholow, B. D., & Anderson, C. A. (2002). Effects of violent video games on aggressive behavior: Potential sex differences. *Journal of Experimental Social Psychology, 38,* 283–290. doi:10.1006/jesp.2001.1502

Bartholow, B. D., Bushman, B. J., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology, 42,* 532–539. doi:10.1016/j.jesp.2005.08.006

Bartholow, B. D., Sestir, M. A., & Davis, E. B. (2005). Correlates and consequences of exposure to video game violence: Hostile personality, empathy, and aggressive behavior. *Personality and Social Psychology Bulletin, 31,* 1573–1586. doi:10.1177/0146167205277205

Beier, R., & Kutzner, F. (2012). *Choose a juice! The effect of choice options and intention on aggression in a modified hot-sauce paradigm*. Paper presented at the 13th annual meeting of Society for Personality and Social Psychology, San Diego, CA.

Ben-Porath, D. D., & Taylor, S. P. (2002). The effects of diazepam (valium) and aggressive disposition on human aggression. *Addictive Behaviors, 27,* 167–177. doi:10.1016/S0306-4603(00)00175-1

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin, 106,* 59–73. doi:10.1037/0033-2909.106.1.59

Bethesda Game Studios. (2006). *The Elder Scrolls IV: Oblivion* [Computer game]. Novato, CA: 2K Games.

Bond, A., Curran, H. V., Bruce, M. S., O'Sullivan, G., & Shine, P. (1995). Behavioural aggression in panic disorder after 8 weeks' treatment with alprazolam. *Journal of Affective Disorders, 35,* 117–123. doi:10.1016/0165-0327(95)00053-4

Bond, A., & Lader, M. (1986). The relationship between induced behavioural aggression and mood after the consumption of two doses of alcohol. *British Journal of Addiction, 81,* 65–75. doi:10.1111/j.1360-0443.1986.tb00296.x

Bond, A., & Lader, M. (1988). Differential effects of oxazepam and lorazepam on aggressive responding. *Psychopharmacology, 95,* 369–373. doi:10.1007/BF00181949

Breuer, J., Scharkow, M., & Quandt, T. (in press). Sore losers? A reexamination of the frustration-aggression hypothesis for colocated video game play. *Psychology of Popular Media Culture*.

Bushman, B. J. (1995). Moderating role of trait aggressiveness in the effects of violent media on aggression. *Journal of Personality and Social Psychology, 69,* 950–960. doi:10.1037/0022-3514.69.5.950

Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public: Scientific facts versus media misinformation. *American Psychologist, 56,* 477–489. doi:10.1037/0003-066X.56.6-7.477

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75,* 219–229. doi:10.1037/0022-3514.75.1.219

Bushman, B. J., & Gibson, B. (2011). Violent video games cause an increase in aggression long after the game has been turned off. *Social Psychological and Personality Science, 2,* 29–32. doi:10.1177/1948550610379506

Bushman, B. J., Ridge, R. D., Das, E., Busath, G. L., & Key, C. W. (2007). When God sanctions killing: Effect of scriptural violence on aggression. *Psychological Science, 18,* 204–207. doi:10.1111/j.1467-9280.2007.01873.x

Carnagey, N. L., & Anderson, C. A. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition, and behavior. *Psychological Science, 16,* 882–889. doi:10.1111/j.1467-9280.2005.01632.x

EA Canada. (2010). *2010 FIFA World Cup South Africa* [Video game]. Redwood City, CA: EA Sports.

Elson, M., & Bente, G. (2009). *CORTEX: Computer-aided registration tool for experiments*. Cologne, Germany: University of Cologne. Retrieved from http://cortex.uni-koeln.de

Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (in press). Comparing apples and oranges? Evidence for pace of action as a confound in research on digital games and aggression. *Psychology of Popular Media Culture*. Advance online publication. doi:10.1037/ppm0000010

Engelhardt, C. R., Bartholow, B. D., & Saults, J. S. (2011). Violent and nonviolent video games differentially affect physical aggression for individuals high vs. low in dispositional anger. *Aggressive Behavior, 37,* 539–546. doi:10.1002/ab.20411

Epic Games. (2007). *Unreal Tournament 3* [Computer game]. Chicago, IL: Midway Games.

Ferguson, C. J. (2007). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior, 12,* 470–482. doi:10.1016/j.avb.2007.01.001

Ferguson, C. J. (2011). The wild west of assessment. Measuring aggression and violence in video games. In L. Annetta & S. C. Bronack (Eds.), *Serious educational game assessment* (pp. 43–56). Rotterdam, the Netherlands: Sense. doi:10.1007/978-94-6091-329-7_3

Ferguson, C. J., & Rueda, S. M. (2009). Examining the validity of the modified Taylor competitive reaction time test of aggression. *Journal of Experimental Criminology, 5,* 121–137. doi:10.1007/s11292-009-9069-5

Ferguson, C. J., Smith, S. M., Miller-Stratton, H., Fritz, S., & Heinrich, E. (2008). Aggression in the laboratory: Problems with the validity of the modified Taylor competitive reaction time test as a measure of aggression in media violence studies. *Journal of Aggression, Maltreatment & Trauma, 17,* 118–132. doi:10.1080/10926770802250678

Giancola, P. R., & Chermack, S. T. (1998). Construct validity of laboratory aggression paradigms: A response to Tedeschi and Quigley (1996). *Aggression and Violent Behavior, 3,* 237–253. doi:10.1016/S1359-1789(97)00004-9

Giancola, P. R., & Zeichner, A. (1995a). Construct validity of a competitive reaction-time aggression paradigm. *Aggressive Behavior, 21,* 199–204. doi:10.1002/1098-2337(1995)21:3<199::AID-AB2480210303>3.0.CO;2-Q

Giancola, P. R., & Zeichner, A. (1995b). An investigation of gender differences in alcohol-related aggression is strongly related. *Journal of Studies on Alcohol, 56,* 573–579.

Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology, 98,* 211–221. doi:10.1037/a0016997

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2,* e124. doi:10.1371/journal.pmed.0020124

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23,* 524–532. doi:10.1177/0956797611430953

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS) [Studies with a German version of the "Positive and Negative Affect Schedule" (PANAS)]. *Diagnostica, 42,* 139–156.

Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior, 25,* 331–348. doi:10.1002/(SICI)1098-2337(1999)25:5<331::AID-AB2>3.0.CO;2-1

Lindsay, J. J., & Anderson, C. A. (2000). From antecedent conditions to violent actions: A general affective aggression model. *Personality and Social Psychology Bulletin, 26,* 533–547. doi:10.1177/0146167200267002

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of *p* values just below .05. *The Quarterly Journal of Experimental Psychology, 65,* 2271–2279. doi:10.1080/17470218.2012.711335

McCloskey, M. S., Berman, M. E., Noblett, K. L., & Coccaro, E. F. (2006). Intermittent explosive disorder-integrated research diagnostic criteria: Convergent and discriminant validity. *Journal of Psychiatric Research, 40,* 231–242. doi:10.1016/j.jpsychires.2005.07.004

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7,* 109–117. doi:10.1177/1745691611432343

Mohseni, M. R. (2013). *Virtuelle Nothilfe. Ein Experiment zum Effekt von virtueller Hilfe, Gewalt und Nothilfe auf Hilfe- und Gewaltverhalten* [An experiment on the effect of virtual helping, violence, and emergency assistance on helping and violent behavior.] (Doctoral dissertation). University of Osnabrück, Osnabrück, Germany.

Neurobehavioral Systems. (2010). *Presentation* (Version 14.5) [Computer software]. Albany, CA: Author. Retrieved from http://www.neurobs.com

Pihl, R. O., Young, S. N., Harden, P., Plotnick, S., Chamberlain, B., & Ervin, F. R. (1995). Acute effect of altered tryptophan levels and alcohol on aggression in normal human males. *Psychopharmacology, 119,* 353–360. doi:10.1007/BF02245849

Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior, 31,* 407–419. doi:10.1002/ab.20066

Rushton, B. (2013). Backdooring it: Defense maneuvers around setback. *Illinois Times.* Retrieved from http://www.illinoistimes.com

Savage, J. (2004). Does viewing violent media really cause criminal violence? A methodological review. *Aggression and Violent Behavior, 10,* 99–128. doi:10.1016/j.avb.2003.10.001

Sestir, M. A., & Bartholow, B. D. (2010). Violent and nonviolent video games produce opposing effects on aggressive and prosocial outcomes. *Journal of Experimental Social Psychology, 46,* 934–942. doi:10.1016/j.jesp.2010.06.005

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. doi:10.1177/0956797611417632

Suris, A., Lind, L., Emmett, G., Borman, P. D., Kashner, M., & Barratt, E. S. (2004). Measures of aggressive behavior: Overview of clinical and research instruments. *Aggression and Violent Behavior, 9,* 165–227. doi:10.1016/S1359-1789(03)00012-0

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality, 35,* 297–310. doi:10.1111/j.1467-6494.1967.tb01430.x

Tedeschi, J. T., & Felson, R. B. (1994). *Violence, aggression, and coercive actions*. Washington, DC: American Psychological Association. doi:10.1037/10160-000

Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior, 1,* 163–177. doi:10.1016/1359-1789(95)00014-3

Tedeschi, J. T., & Quigley, B. M. (2000). A further comment on the construct validity of laboratory aggression paradigms: A response to Giancola and Chermack. *Aggression and Violent Behavior, 5,* 127–136. doi:10.1016/S1359-1789(98)00028-7

Valadez, J. J., & Ferguson, C. J. (2012). Just a game after all: Violent video game exposure and time spent playing effects on hostile feelings, depression, and visuospatial cognition. *Computers in Human Behavior, 28,* 608–616. doi:10.1016/j.chb.2011.11.006

Veit, R., Lotze, M., Sewing, S., Missenhardt, H., Gaber, T., & Birbaumer, N. (2010). Aberrant social and cerebral responding in a competitive reaction time paradigm in criminal psychopaths. *NeuroImage, 49,* 3365–3372. doi:10.1016/j.neuroimage.2009.11.040

Verona, E., Reed, A., Curtin, J. J., & Pole, M. (2007). Gender differences in emotional and overt/covert aggressive responses to stress. *Aggressive Behavior, 33,* 261–271. doi:10.1002/ab.20186

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. doi:10.1037/0022-3514.54.6.1063

Weisman, A. M., Berman, M. E., & Taylor, S. P. (1998). Effects of clorazepate, diazepam, and oxazepam on a laboratory measurement of aggression in men. *International Clinical Psychopharmacology, 13,* 183–188. doi:10.1097/00004850-199807000-00005

Zillmann, D. (1983). Arousal and aggression. In R. G. Geen & E. Donnerstein (Eds.), *Aggression: Theoretical and empirical reviews* (Vol. 1, pp. 75–102). New York, NY: Academic Press.

Received March 11, 2013
Revision received November 6, 2013
Accepted November 25, 2013 ■