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

Empirical Evidence, Perspectives, and a Debate Gone Astray

Malte Elson and Christopher J. Ferguson

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
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 psychological 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, Giommetti 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, Bégue, 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, Ulhmann 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 Blumenke, 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. Schmiernbach’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. Ulhmann 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 feel angry). 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, & Scharlow, 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, Ivarsso, & 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, Wittchof, & 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, Vogelsang, 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 Krahe (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 pro-social behaviors 1 year later (Ferguson, Garza, Jerabeck, Ramos, & Galindo, 2013). Instead of the standard self-report measures, von Salisch et al. (2011) used expert rat- ings 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 consequently 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


Hasan, Y., Bègue, L., & Bushman, B. J. (2012). Viewing the world through “blood-red tinted glasses”: The hostile


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.