Page Proof Instructions and Queries

Thank you for choosing to publish with us. This is your final opportunity to ensure your article will be accurate at publication. Please review your proof carefully and respond to the queries using the circled tools in the image below, which are available in Adobe Reader DC* by clicking **Tools** from the top menu, then clicking **Comment**.

Please use *only* the tools circled in the image, as edits via other tools/methods can be lost during file conversion. For comments, questions, or formatting requests, please use ⊤. Please do *not* use comment bubbles/sticky notes 💬.

*If you do not see these tools, please ensure you have opened this file with **Adobe Reader DC**, available for free at get.adobe.com/reader or by going to Help > Check for Updates within other versions of Reader. For more detailed instructions, please see us.sagepub.com/ReaderXProofs.

No.	Query
1	Please (a) check that all authors are listed in the proper order; (b) clarify which part of each author's name is the surname; (c) verify that all author names are correctly spelled/punctuated and are presented in a manner consistent with any prior publications; and (d) check that all author information, such as affiliations and contact information, appears accurately. In particular, please provide or confirm departmental affiliations.
2	Please review the entire document for typographical errors, mathematical errors, and any other necessary corrections; check headings, tables, and figures.
3	Please note that we cannot add additional ORCID iDs to any article at the proof stage. According to ORCID's guidelines, the publisher can include only ORCID iDs that the authors specifically provided for each manuscript before official acceptance for publication.
4	I'm not sure what this means. (And APA says no hyphen after "pseudo.")
5	Per APA style, model names are not capped.
6	Please provide ref citation and add to ref list.
7	Please provide reference citation and add to ref. list.
8	Is this number 4 in the list?
9	Please cite Table 4 in text.

Reexamining the Findings of the American Psychological Association's 2015 Task Force on Violent Media: A Meta-Analysis

Christopher J. Ferguson[1] , Allen Copenhaver[2], and Patrick Markey[3][AQ: 1][AQ: 2][AQ: 3]
[1]Department of Psychology, Stetson University; [2]Criminal Justice, Lindsey Wilson College; and [3]Department of Psychology, Villanova University

Abstract

In 2015, the American Psychological Association (APA) released a task-force technical report on video-game violence with a concurrent resolution statement linking violent games to aggression but not violent crime. The task-force report has proven to be controversial; many scholars have criticized language implying conclusive evidence linking violent games to aggression as well as technical concerns regarding the meta-analysis that formed the basis of the technical report and resolution statement. In the current article, we attempt a reevaluation of the 2015 technical report meta-analysis. The intent of this reevaluation was to examine whether the data foundations behind the APA's resolution on video-game violence were sound. Reproducing the original meta-analysis proved difficult because some studies were included that did not appear to have relevant data, and many other available studies were not included. The current analysis revealed negligible relationships between violent games and aggressive or prosocial behavior, small relationships with aggressive affect and cognitions, and stronger relationships with desensitization. However, effect sizes appeared to be elevated because of non-best-practices and researcher-expectancy effects, particularly for experimental studies. It is concluded that evidence warrants a more cautious interpretation of the effects of violent games on aggression than provided by the APA technical report or resolution statement.

Keywords

violence, aggression, science policy, video games

There has been a great deal of research conducted on whether violent video games cause aggression and violence. Despite the amount of research conducted on the topic, violent-video-game scholars cannot reach consensus as to the true effects of violent video games (Quandt et al., 2015). Many scholars continue to assert that violent video games cause real-world aggression and subsequent violence, whereas other scholars dismiss such claims, citing methodological issues (e.g., Sauer, Drummond, & Nova, 2015) and publication bias (i.e., Segev et al., 2016).

In an attempt to provide a declarative summary of violent-video-game research, the American Psychological Association (APA) established its Task Force on Violent Media in 2013. The task force conducted a meta-analysis of a small number of existing studies on the associations of violent video games with aggression and prosocial outcomes. The task force concluded that violent video games were associated with aggressive outcomes and a lack of prosocial outcomes (APA, 2015b), albeit evidence was lacking for associations with violent crime. Both the process and findings of the APA's task force have been questioned by violent-video-game scholars for a variety of reasons (e.g., Copenhaver & Ferguson, 2018; Kühn et al., 2018b; Quintero-Johnson, Banks, Bowman, Carveth, & Lachlan, 2014). These criticisms include but are not limited to the composition of

Corresponding Author:
Christopher J. Ferguson, Department of Psychology, Stetson University, DeLand, FL 32723
E-mail: cjferguson1111@aol.com

the members of the task force, methodological issues, and conclusions drawn from the meta-analysis. Thus, the purpose of this study is to conduct an independent evaluation of the 2015 APA meta-analysis of the research on violent video games and their associations (if any) to aggression and prosocial outcomes. It is the purpose of this reevaluation effort to examine whether the APA's 2015 resolution on video-game violence is based on a sound data foundation and whether claims made about the field of video-game violence are accurate. This reevaluation thus concerns itself mainly with the state of the science in 2015, when these conclusions were made by the APA, and is not an update on the current state of the science. It would be reasonable to suggest that the science might have changed since 2015, which would not necessarily mean the APA's conclusions in 2015 were unsound. However, if the data were unable to support any effects of violent games on aggression, the APA's 2015 task-force statement and resultant resolution would be called into question.

Violent-Video-Game Controversy and Research

Violent video games have a history of being blamed for mass shootings and school shootings in particular. After a mass shooting, violent video games are frequently cited as the cause of the shooting by media, politicians, and moral advocates (Sternheimer, 2007; Zgoba, 2005). From the 1999 Columbine High School shooting (in which two boys killed 12 students and a teacher before killing themselves) to the more recent 2018 Parkland, Florida, shooting (in which a 19-year-old former student killed 17 at a high school), violent video games have been criticized in news media as mediums by which children are trained to kill (i.e., S. Olson, 2018) and become desensitized to violence (i.e., Ducharme, 2018).

Politicians have been particularly critical of the supposed effects violent video games have on children. Arguably the most notable is that after the shooting on Valentine's Day 2018 in Parkland, Florida, President Donald Trump blamed violent video games. His comments resulted in a meeting with representatives of the video-game industry at the White House; this meeting was not dissimilar to the 2013 meeting Vice President Joe Biden had with industry leaders after the December 2012 shooting at Sandy Hook Elementary School in Newtown, Connecticut, which was perpetrated by a 20-year-old man (M. Anderson, 2018). These actions are nothing new; former President Bill Clinton blamed the video games Mortal Kombat, Killer Instinct, and Doom for the 1999 Columbine shooting (Gerstenzang, 1999). Such has been the trend of political response to the perceived threats of violent games since Senator

Joe Lieberman claimed in 1993 that Mortal Kombat taught kids to enjoy inflicting gruesome cruelty (Crossley, 2014).

Politicians have attempted a variety of strategies to deal with the violent-video-game problem. Legislators at the state and federal levels have proposed a number of bills designed to curtail the alleged impact that violent video games have on American youths. Those measures include but are not limited to increased taxes on the sale of violent video games, placing aggression warning labels on their packaging, and preventing minors from purchasing them (Copenhaver, 2015).

The most noteworthy case of attempting to address the violent-video-game problem comes from a 2005 California law that aimed to control the purchase of violent video games by minors, which subsequently necessitated review by the Supreme Court of the United States (SCOTUS). In this case, California had created legislation that would give the state the ability to rate video-game content and impose fines on distributors that sold highly violent games to minors. This legislation was quickly struck down by the lower courts as unconstitutional and lacking in research support. In 2011, the SCOTUS ruled in *Brown v. Entertainment Merchants Association* that limiting the sale of violent video games to minors is unconstitutional because such a law violates free-speech protections, affirming the lower courts. In addition, the SCOTUS criticized the research California relied on in its argument that violent video games cause aggression. The SCOTUS disparaged such research for poor operationalization of study variables and a reliance on research evidence that failed to demonstrate a causal link between violent video games and aggression (Ferguson, 2013). In other words, various findings on the effects of violent games on players provide too much question as to the true effects of violent games on players, which makes the application of such questionable associations difficult for lawmakers and policymakers. Therefore, as a result of the decision, the SCOTUS established precedent that violent-video-game sales to children cannot be regulated. However, Copenhaver (2015) found in the wake of the Sandy Hook shooting that politicians engaged in proposing pseudo-agenda **[AQ: 4]**bills very similar to the law the SCOTUS had already deemed unconstitutional.

There is a great deal of psychological and criminological research on the effects of violent video games on players. However, no consensus exists among scholars as to whether violent-video-game exposure negatively affects players (J. D. Ivory et al., 2015). The central debate about the effects of violent video games revolves around whether violent video games correlate with aggressive behavior. The first important study in this area was C. A. Anderson and Bushman's (2001)

meta-analysis of the effects of violent video games on aggression for children and young adults alike. They found that violent video games increase aggressive behavior, increase physiological arousal and aggression-related feelings and thoughts, and diminish behaviors that are prosocial in nature. Bushman and Anderson (2002) then found that players of violent video games have more aggressive thoughts and feel angrier than players of nonviolent games. Bushman and Anderson argued that consistent with the general-aggression model, **[AQ: 5]**such game play provided aggressive scripts and schemata and caused players to adopt an aggressive response to ambiguous behavioral stimuli. C. A. Anderson (2004) later provided an updated meta-analysis on the effects of violent games on aggression and found that, on the basis of experimental studies, the relationship is *causal*. Today, some researchers and professional agencies still consider violent video games to be a risk factor for aggression (see APA, 2015a; Calvert et al., 2017; Scharrer, Kamau, Warren, & Zhang, 2018).

Research also exists that demonstrates that violent video games are not linked to aggression either correlationally or causally. For example, Ferguson, Barr, et al. (2015) found in two experimental studies and one correlational study of youths that violent video games did not increase youth aggression and that there is no correlation between gaming and youth aggression. Numerous other recent studies have come to similar conclusions correlationally (e.g., DeCamp, 2015; Przybylski & Weinstein, 2019b), longitudinally (Ferguson & Wang, 2019; Kühn et al., 2018b; Lobel, Engels, Stone, Burk, & Granic, 2017), and experimentally (Devilly, Brown, Pickert, & O'Donohue, 2017; Hilgard, Engelhardt, Rouder, Segert, & Bartholow, 2019). Although some scholars may argue that the interpretation of null results is difficult, failing to consider null results obviously renders a hypothesis unfalsifiable. Furthermore, new techniques involving rigorous examination of effect sizes (Przybylski & Weinstein, 2019a) or Bayesian analyses (Hilgard et al., 2019) can improve the ability to find support for the null.

Research on the potential harmful effects of violent video games expanded beyond experimental and correlational studies as it pertained to aggression and began to include studies on associations between violent-video-game play and real-life violence. As with the aggression studies, the results of these studies are mixed. For example, DeLisi, Vaughn, Gentile, Anderson, and Shook (2013) found violent-video-game play was associated with violent delinquency in a sample of institutionalized juvenile delinquents. In addition, Greitemeyer and Sagioglou (2017) found an association between violent-video-game exposure and sadism. However, other researchers have called into question the strength of the results of such studies by indicating these associations vanish after other variables, such as gender, family environment, or mental health, are controlled (see Gunter & Daly, 2012); other researchers have failed to find evidence that such a relationship exists (see Breuer, Vogelgesang, Quandt, & Festl, 2015; DeCamp & Ferguson, 2017; Przybylski, Deci, Rigby, & Ryan, 2014). A time-series-analysis study conducted by Markey, Markey, and French (2015) even found exposure to violent video games to be associated with decreases in real-world violence, and various other studies have replicated this finding (e.g., Beerthuizen, Weijters, & van der Laan, 2017; Cunningham, Engelstätter, & Ward, 2016).

What Do Best-Practices Studies Look Like?

One ongoing controversy with meta-analysis is the "garbage in, garbage out" (GIGO) phenomenon, which can cause meta-analyses to significantly misinterpret the strength of evidence for a hypothesis. Essentially, GIGO notes that meta-analyses are only as good as the included studies, and when included studies have systematic methodological problems (and this concern has been expressed for violent-video-game research), meta-analyses that do not take these into consideration can cause more confusion than illumination. Thus, it is worth considering what best-practices studies look like in violent-video-game research. We consider experimental and correlational and longitudinal studies separately.

For experimental studies, participants should undergo random assignment to video-game conditions (although this would seem obvious, this has not always been the case). Violent- and nonviolent-video-game conditions should include games that are as close as possible in other qualities (storyline, characterization, difficulty, competitiveness, etc.) aside from violent content. This is the only way to isolate violent content as the variable of interest for any observed differences. Other variables such as difficulty of the game (Przybylski et al., 2014), competitiveness (Adachi & Willoughby, 2011a), or pace of action (Elson, Breuer, Van Looy, Kneer, & Quandt, 2015) have been shown to increase aggression; thus, it is essential that different game conditions control for these variables. This is also why a no-game control condition is ineffective because such studies leave numerous variables uncontrolled to the extent that these should not be considered violent-video-game studies at all. In effect, when comparing violent games to nothing at all, it is very difficult to conceptualize what is actually being compared. Such a basic level of comparison is viewed as essential to scholars on both sides of this debate (e.g., C. A. Anderson, 2004). Even when violent and nonviolent games are compared, often other variables, such as those mentioned, are poorly

controlled (Adachi & Willoughby, 2011a). Aggression measures should be standardized to cut down on researcher-expectancy effects (Elson, Mohseni, Breuer, Scharkow, & Quandt, 2014). Unfortunately, standardization of outcome measures has not been normative until more recent years (e.g., after 2013).

For correlational and longitudinal studies, best-practices studies control for theoretically relevant third variables (i.e., gender, age, personality family environment, mental health history, peer environment, and for longitudinal studies, Time 1 aggression). Standardized aggression measures should be used, as with experimental studies; the ideal is using clinically validated scales such as the Child Behavior Checklist. [AQ: 6] Video-game violence should be assessed using independent ratings either by scholars or by ratings boards (e.g., Entertainment Software Ratings Board, Pan European Game Information, etc.) rather than respondent self-report because self-report of violent content tends to alert respondents of study hypotheses and potentially introduces demand characteristics.

Thus, best-practices studies in both realms make clear efforts to isolate the variable of interest (i.e., violent content) from other confounding variables. Furthermore, best-practices studies in both realms employ a high degree of measurement standardization, which reduces researcher-expectancy effects. When possible, best-practices studies also take efforts to reduce demand characteristics or assess for participant suspiciousness of the hypotheses.

Non-best-practices (NBP) studies, by contrast, allow confounding variables to intrude or do not control for researcher-expectancy issues, which potentially influence effect sizes. In some cases, such as failure to compare violent games with nonviolent games at all (even if imperfectly; e.g., comparing different violent games to each other, or violent games to violent movies, or using a no-game control), the results may be so conceptually different from a best-practices study that they do not exist within the realm of appropriate studies for inclusion at all. Thus, we considered three tiers of quality: (a) violent-game studies that use best practices, (b) violent-game studies that do not use best practices, and (c) studies that, because of serious basic problems, are not truly violent-video-game studies at all.

Controversy Surrounding the 2015 APA Task Force on Violent Media

In 2013, the APA assembled a task force to study the alleged associations among violent video games and aggression and prosocial outcomes. With the creation of the 2015 Task Force on Violent Media, the APA aimed to update its 2005 *Resolution on Violence in Video*

Games and Interactive Media. The task-force members convened, selected studies for inclusion in the meta-analysis, and ultimately decided violent video games increase aggression and decrease prosocial behavior (APA, 2015b). Although dozens of studies were available during the time frame the APA task force considered, only 18 were included in the meta-analysis that provided the foundation for the APA's conclusions. The APA went so far as to release a press release stating, "APA Review Confirms Link Between Playing Violent Video Games and Aggression" (APA, 2015a). With its findings, the task force also made a variety of policy recommendations, arguably the most important being the call for future governmental studies on effects of violent games and an update to the ratings system of the Entertainment and Software Ratings Board.

It is worth examining whether the APA task force and ensuing 2015 resolution statement intended to causally associate violent games with increased aggression and other outcomes. The task-force report (APA, 2015b) often uses the language of association. However, authors of the report also made it clear that they were making causal inferences in the section titled "Causality and ethical conduct of research" (APA, 2015b, p. 14), in which they asserted that although no one method of study is perfect, the confluence of data (in the eyes of the task force) allowed for causal inferences. That section was concluded with the clearly causal statement, "On the basis of the body of empirical evidence, we concluded that the impact of exposure to violent video game use on aggressive outcomes is robust" (APA, 2015b, p. 15). Likewise, although the APA resolution often used "association" language, authors of the resolution mentioned "the effects of violent video game use" multiple time and suggested interventions, both of which assert causality. Even if we were to generously interpret association language as correlational, it is reasonable to expect that most policymakers, the public, and indeed, even scientists would interpret such statements as causal. Furthermore, in defending the resolution and task-force report in news media, the task-force chair, Mark Appelbaum, compared the impact of video games on aggression with the impact of taking aspirin on heart disease, a clearly causal inference (e.g., Wofford, 2015).

With the APA's announcement of the creation of a task force to study the effects of violent video games, 230 video-game scholars, criminologists, and psychologists came together to sign a unified statement asking the APA to no longer make "state of knowledge"/policy statements related to this field (see Consortium of Scholars, 2013). These scholars and others were concerned with a variety of problems inherent to both the creation of the task force and how it carried out its work. First, scholars contended the committee was

created by selecting academics sympathetic to the cause of addressing perceived social ills stemming from violent-video-game play (Wofford, 2015). This critique was raised because four of the seven task-force members had taken public antigame stances before being named to the task force, and no task-force members had taken progame stances. Two members (Hamby, Dodge) had signed an amicus brief supporting the regulation of violent games in the *Brown v Entertainment Merchants Association* (2011) Supreme Court case discussed below. One had signed onto another public statement linking violent media, including games, to societal violence (Graham). And a fourth (Calvert) had specifically done research on violent games despite the task force being set up to specifically exclude individuals with a professional stake in the topic (Wofford, 2015). Given how many APA members exist who had never taken a public position on games, this arrangement appeared to be statistically improbable as a chance occurrence. In addition, the meta-analysis was critiqued for its lack of inclusion of studies that actually met the task force's own stated inclusion criteria (Wofford, 2015). The task force was also criticized largely for failing to conduct a meta-analysis study in an impartial and inclusive fashion absent the methodological concerns that have hindered much of the violent-video-game research in the field (see Elson et al., 2014). It was also unclear whether the 18 studies included in the 2015 task-force meta-analysis were representative of the field.

The Current Study

Given the controversies that have followed the APA's task force and the resultant resolution statement, it is worth taking another look at whether the conclusions of the APA (2015b) were well supported by the data available at the time. In the current article, we provide a reevaluation effort of the meta-analysis that provided the bedrock for the APA's 2015 resolution on violent video games. As noted above, it was the purpose of this reevaluation to examine whether the conclusions of the task force and resultant APA resolution were founded in sound data and an accurate representation of the field as it existed at the time of the resolution in 2015. By contrast, it was not the purpose of this analysis to provide an updated summary of the research conducted since 2015. In particular, this reevaluation effort is concerned with several related issues:

1. Did the 2015 task force include a fully representative sample of articles on effects of violent games?
2. Was the overall effect size provided by the 2015 task force able to represent a population-level effect of violent video games on aggression?

3. Would conclusions about effects of violent games on aggression have changed had the task force considered methodological critiques that have raged in the field for years. In other words, is there concern that the conclusions of the 2015 APA task force and the APA's subsequent resolution on violent video games were the product of systematic methodological flaws causing false positive results?

Method

Criteria for article inclusion in the 2015 meta-analyses

To fully understand the 2015 task-force meta-analyses, data on study inclusion and effect size were sought directly from the APA. We are grateful to the APA for providing this information. This allowed us to examine both effect sizes and decisions regarding which studies were included in the 2015 meta-analysis.

The 2015 task force narrowed down an original pool of 170 studies to 31 on the basis of a set of criteria. The task-force technical report listed several criteria used for study inclusion in the meta-analysis (APA, 2015b, p. 8). From the task-force technical report, the winnowing process appears to have included criteria related to "causal inference, ecological validity, sampling validity, and measurement of independent and dependent variables" (p. 8). However, none of these criteria were explained in detail (e.g., how correlational studies were included despite the fact that causal inference was one criterion.) The summary of these criteria in the technical report is as follows:

> The studies were then divided into two groups: studies perceived as having sufficient utility and those perceived as having insufficient utility for informing the decisions and recommendations the task force was charged with making. Studies were assigned to the sufficient utility group if they were rated by at least one rater as having sufficient ecological validity, sampling validity, or possibility for causal inference to address the task force's charge and no more than one of these variables was rated as having insufficient utility by the second rater. In addition, to be included in the sufficient utility group, the study had to have at least one dependent and one independent variable rated as having sufficient measurement validity. (APA, 2015b, p. 6)

No further details on how these determinations were made were provided.

Measures or surveys used in studies were evaluated on validity, reliability, and precision (also not clearly defined). For this latter evaluation, the authors wrote only,

> We used three primary criteria for evaluating the measures in a study: validity, reliability, and precision. Studies that ranked high on at least one of these factors and low on none of the factors were rated as having sufficient utility. Studies that did not meet these criteria were assigned to the insufficient utility group. Interrater reliability was high, and any initial differences of more than one unit were resolved by having both raters rescore the article. (APA, 2015b, p. 6)

No details on how such evaluations were made were provided, which appears to be a critical oversight, particularly given considerable controversies regarding the validity, reliability, and precision of many instruments used in this field (McCarthy & Elson, 2018).

With that in mind, it is also unclear whether the task force applied its criteria in a reliable way. For instance, some studies that used clinically validated measures such as the Child Behavior Checklist were excluded from the analysis, yet the task force included studies employing the Competitive Reaction Time Test,[AQ: 7] a measure repeatedly criticized for years exactly for its lack of validity, reliability, and precision (McCarthy & Elson, 2018; Ritter & Eslea, 2005; Savage, 2004; Tedeschi & Quigley, 1996). In some cases, studies that appear to have attempted to improve on the problems of poor aggression measures were rejected from the 2015 meta-analysis (e.g., Adachi & Willoughby, 2011a; Tear & Nielsen, 2013). Without a clearer understanding of how such determinations were made, the reliability of how the criteria were applied appears uncertain. Even if we were to credit that most aggression measures have significant problems, the pattern of inclusion and exclusion of studies by the 2015 task force was difficult to discern and replicate.

Reproducibility of the 2015 meta-analysis

The task force included only 18 studies in its final report, citing difficulties calculating effect sizes for the remainder. In our examination of the results of these evaluations, we found them difficult to reproduce. In the introduction, we discussed the qualities of best-practices studies as well as some issues we identified as potentially disqualifying studies from the realm of appropriate inclusion entirely. For instance, five studies included among the 18 (28%) had significant limitations that should have violated those inclusion criteria. These

studies included the following: (a) Happ, Melzer, and Steffgen (2013) contrasted different characters in a violent video game but had no nonviolent-game control; (b) Krcmar and Lachlan (2009) failed to include a non-violent-game control condition and used a substandard no-game control instead; (c) Krcmar and Farrar (2009) failed to include a nonviolent-game control condition and used a substandard no-game control instead; (d) Lin (2013) merely compared movies and video games with no nonviolent controls; and (e) Wang et al. (2009) did not include any measures of aggressive behavior and relied on functional MRI (fMRI) results only, the interpretation of which can be ambiguous (Kühn et al., 2018b). The lack of a nonviolent-game control is a critical one. Without a carefully matched nonviolent-game control, it is impossible to isolate violent content as the essential component of any change in aggression. Even when a nonviolent game is used as a control, it is essential that the nonviolent game be carefully matched to the violent game as closely as possible on characteristics other than violent content (Adachi & Willoughby, 2011b; Elson & Quandt, 2016). Indeed, an increasing array of studies suggests that when games are cautiously matched, it is not violent content but other characteristics such as competitiveness (Adachi & Willoughby, 2011a), pace of action (Elson et al., 2015), difficulty (Kneer, Elson, & Knapp, 2016), and frustration (Przybylski et al., 2014) that are related to aggression.

Thus, the failure to provide a matched-nonviolent-control game violates the task force's own criteria regarding both causal inference and measurement of the independent variable given that the independent variable of interest (violent-video-game content) simply was not isolated and thus provides no helpful information (Elson & Quandt, 2016). Regarding the fMRI study (Wang et al., 2009), another similar study with fMRI results (Regenbogen, Herrmann, & Fehr, 2010) was rejected by the task force, and we could discern no reason for this discrepancy given the task force's own criteria. fMRI results can also be notoriously difficult to interpret when not properly keyed to outcome behaviors (e.g., Vul, Harris, Winkielman, & Pashler, 2009), once again undercutting the inclusion criteria of causal inferences and, in this case, measurement of the dependent variable. We inquired of several task-force members, including the chair, about how effect sizes had been calculated from these studies as well as to help us understand the inclusion and exclusion criteria they used. Despite repeated efforts to seek clarification on these important issues, we never received any substantial reply to our inquiries. Given that none of these studies provided data we felt to be sufficient for calculation of a meaningful effect size, they were not included in our reevaluation project.

Given our difficulty in reproducing a similar set of studies using the 2015 task force's inclusion criteria, it appeared prudent to attempt a reevaluation from the ground up, including new study search, inclusion criteria, best practices, and effect size extraction. What follows is our procedure for reproducing the 2015 meta-analysis.

Selection of studies

Identification of relevant studies involved a search of the PsycINFO, MEDLINE, and Digital Dissertations databases using the search terms Violence AND ("video game*" or videogame* or "computer game*") AND (aggress* OR prosocial OR empathy OR desensitization OR antisocial) as subject searches. Consistent with the time frame for the 2015 meta-analysis, the years 2009 to 2013 were searched. Unpublished studies, aside from dissertations, were not considered given that such searches tend to produce nonrepresentative biased samples of unpublished studies that add bias to meta-analyses rather than eliminating bias (Ferguson & Brannick, 2012) and do little to address publication-bias issues (Tsuji, Christia, Frank, & Bergmann, 2019). This decision appeared consistent with the APA task force, which also did not include unpublished studies in its meta-analysis. Included studies had to meet the following criteria:

1. Each study had to measure the influence of video games, whether violent or nonviolent, on at least one of the outcomes related to aggression, whether it be behavior, affect, cognition, prosocial behavior, or desensitization. Studies that did not distinguish video games from other media were not included. We were interested in obtaining a broad but relevant sample, and thus, included studies had to fit one of the following three criteria:
   a. experimental studies comparing a violent game with a nonviolent control game using random assignment with an outcome measuring aggression (behavior, affect, or cognition), prosocial behavior, or desensitization (dehumanization, reduced empathy, or emotional response to violence);
   b. correlational studies measuring exposure to violent game content specifically (i.e., not general game playing) with an outcome measuring aggression, prosocial behavior, or desensitization;
   c. longitudinal studies measuring exposure to violent game content specifically (i.e., not general game playing) with an outcome measuring aggression, prosocial behavior, or desensitization.

2. Each study had to present statistical outcomes or data that could be meaningfully converted into an effect size (r).
3. Experimental studies had to contrast violent-video-game play with nonviolent-video-game play. Studies that did not include a control condition consisting of a nonviolent video game or games were not included, nor were studies primarily examining media-literacy interventions or contrasting playing as opposed to watching video games. Studies that used nongame controls were likewise excluded given that absence of any game play introduces confounds that threaten internal validity.
4. A given sample was included only once in the meta-analyses to maintain independence. Some samples, including longitudinal studies, may produce multiple publications, but only one such study was included in the current analysis.[1] Decisions were made in effect-size selection in preference for longer longitudinal periods as well as better controlled/more conservative analyses.
5. Studies that failed to include a dependent variable directly measuring one of the key outcomes (aggression, prosocial behavior, desensitization) were considered to lack ecological validity. This was most relevant to some fMRI studies—those that assume a brain difference is related to some outcome such as aggression—without actually measuring aggression.

Other inclusion criteria used by the 2015 task force, such as validity of measures, were included as moderator variables. Rather than excluding studies, this allowed us to examine how best practices or NBPs may influence effect sizes, an issue that was not considered by the 2015 task force. These are discussed in the Moderator Analyses section. We did not use an inclusion criterion related to causal inferences because this would not apply to correlational and longitudinal studies. Likewise, almost all samples are convenience samples, so we were unsure of how to apply a "sampling validity" criteria (unless, e.g., an experiment used nonrandom assignment).

The initial search (carried out in June 2018) returned approximately 182 hits, many of which were either nonempirical or otherwise did not meet the inclusion criteria above. Eliminating those studies resulted in 68 articles originally identified by the 2015 task force as well as 29 additional articles that had not been identified by the 2015 task force. Employing the inclusion criteria, the final search included 13 of the original 18 studies included in the 2015 task force (five were rejected for reasons explained above, mainly involving the lack of data available to calculate a meaningful effect size), 30

studies rejected by the 2015 task force that met our inclusion criteria, and finally, 18 new studies that had not been considered by the 2015 task force.

We felt it important to run a separate search for literature given that it was possible the task force had missed some articles. Although the task force provided search terms in its technical report (APA, 2015b), it was not clear how the search terms were used (i.e., in what combination, or whether using *subject* searches, etc.). As a consequence, we found 18 additional articles during the task-force time frame that were not found by the original task force, which suggests that this new search was a valuable source of information. In addition, we are uncertain why the APA task force initially rejected the 30 articles, given that they appear to meet the task force's inclusion criteria. Personal communication with the task force members did not result in further elucidation. For the current analysis, each article was assessed by two raters, each blinded to the other's ratings for inclusion. Krippendorf's α reliability on the inclusion decision was .78; discrepancies were then resolved by consensus of all researchers. This process was completed by September 2018. The list of studies along with effect-size estimates is available at https://osf.io/fpua4/. Details on data extracted from each article are described below under effect-size estimates and moderator analyses. A full list of studies broken down by (a) those appearing in the original 2015 task force report, (b) those appearing in the original 2015 task force report from which we could replicate effect sizes, and (c) those included in our final meta-analysis is provided in the Supplemental Material available online. A PRISMA chart is provided as Appendix Figure A1.

Effect-size estimates

In the current article, we made use of effect sizes in the metric of *r*. When possible, particularly for correlational and longitudinal studies, we used results that are based on multivariate analyses resulting in standardized regression coefficients (βs). Reasons for a preference for βs in meta-analysis are numerous, primarily given the concern that bivariate *r* may return spuriously high effect-size estimates that do not reflect real correlations after important factors are controlled (Pratt et al., 2010; Savage & Yancey, 2008). Use of βs makes more sense theoretically given that most multivariate analyses include theoretically relevant controls.

Many meta-analyses in prior years had relied on bivariate *r* in the hopes that using *r* rather than βs would result in more homogeneous analyses. Increasingly, in recent years, debate has focused on whether bivariate *r*s or standardized regression coefficients are more appropriate to use in meta-analyses, particularly

with correlational or longitudinal studies. For instance, some scholars have noted that inferring bivariate *r*s using βs when the goal is to meta-analyze bivariate *r*s can be problematic (Roth, Le, Oh, Van Iddekinge, & Bobko, 2018). However, this applies to situations in which βs are being used to estimate bivariate *r* when the bivariate *r* is not reported in an article and the goal is to conduct a meta-analysis of bivariate *r*s. This does not necessarily extend to meta-analyses that are specifically focused on βs.

Of central interest to the current article is not whether βs should be used to infer bivariate *r*s when bivariate *r*s are missing but rather whether βs should be used as the central metric in meta-analysis. As noted above, there are strong theoretical reasons for doing so given that bivariate *r*s are likely to overestimate true population effect sizes. However, debate does exist on whether the use of βs is methodologically appropriate. This debate can be seen in regard to another recent meta-analysis in this field that used standardized regression coefficients (Ferguson, 2015a). Rothstein and Bushman (2015), commenting on Ferguson (2015a), claimed that βs have many problems because βs create different, potentially higher variances and different distributions than bivariate *r*s, with bivariate *r*s more appropriate for use with meta-analysis.

In a reply to this comment, Ferguson (2015b) examined the distributions and variance of βs and bivariate *r*s for the pool of studies available at the time in effects of violent games on children and discovered that the variance and distribution of effect sizes for βs was *better* suited to meta-analysis than were bivariate *r*s. This is likely because the presumed/ideal distribution and variance of bivariate *r*s is far different from the actual variance and distribution of bivariate *r*s given that most studies vary widely in measurement, analytics, and sample. In a later response to Rothstein and Bushman (2015), Furuya-Kanamori and Doi (2016) independently examined their critiques. Furuya-Kanamori and Doi noted that βs produce a closer estimate of underlying effect size than do bivariate *r*s and that, contrary to the claims of Rothstein and Bushman, do not have greater variances than bivariate *r*. Using Monte Carlo simulation, Furuya-Kanamori and Doi confirmed that βs are appropriate for use in meta-analysis and do not produce erroneous effect-size estimates. Note that Furuya-Kanamori and Doi were able to independently reproduce the Ferguson (2015a) meta-analysis. Since this debate, other meta-analyses in video-game research (e.g., Mathur & VanderWeele, 2019; Prescott, Sargent, & Hull, 2018) have made use of βs in meta-analysis, often comparing them with the results from bivariate *r*s. This suggests that such an approach has become increasingly conventional for this field, and we adopt it here.

In other realms, the adoption of βs in meta-analyses has become increasingly common. As noted above, they have been strongly advocated among criminologists (Pratt et al., 2010; Savage & Yancey, 2008), researchers in a field that bears some tangential relationship to the issues under consideration here. Concerns do emerge that model-specification variance among studies may introduce heterogeneity (Lipsey & Wilson, 2001), and we are mindful of these concerns. However, as noted above, it appears that it was assumed bivariate *r*s would not have similar problems, although apparently they do and to greater degrees (albeit for reasons other than model specification). Other scholars have concluded that, despite these concerns, βs are appropriate for use in meta-analysis and have superior qualities to bivariate *r* (Bowman, 2012).

Ultimately, we were persuaded by arguments that βs can be meaningfully used in meta-analysis. However, we also understand debate is likely to continue. Because it has become common practice for meta-analyses in media effects to present outcomes for both βs and bivariate *r*s, for correlational and longitudinal studies, we also present both. This will allow readers to come to their own conclusions about the meaningfulness of the difference between them.

Thus, in this study, we employed βs as effect-size estimates for correlational and longitudinal studies. All included articles provided β estimates, and βs were not pooled alongside bivariate *r*s. In the case of experimental results, effect sizes were computed directly from group means and standard deviations or, if not available, *t* and *F* statistics.

In cases in which articles presented more than one effect-size estimate, they were aggregated for an average effect size. Generally, for the included studies, when multiple outcomes were used, heterogeneity in effect sizes was low, which suggests that aggregation was appropriate. Given that such measures were typically of the same construct, assumption of high correlation between conceptually similar outcomes warrants simple aggregation (Pustejovsky, 2019). To double-check this assumption, we examined a subset of effect sizes using the Mad aggregation function based on the Borenstein, Hedges, Higgins, and Rothstein (BHHR) technique, which adjusts effect sizes for presumed intercorrelation between measures or outcomes (Del Re & Hoyt, 2019). Results using this method were no different from raw aggregates, so raw aggregates were retained throughout. Some articles presented multiple competing statistical models with different effect-size estimates, particularly for multivariate analyses. When this occurred, the most conservative model (e.g., that which employed theoretically relevant controls related to gender, prior aggression, family environment, and

mental health) was used as the effect-size estimate for the controlled analyses.

Moderator analyses

Several moderators were considered as potentially important for the current article. These moderators included some issues that original task force members had listed as inclusion criteria (albeit unreliably), such as the validity of measures. This allows for a more considered examination of how different best practices or NBPs can influence effect sizes. Study year was considered as a moderator along with the age of participants. Studies were also coded for best practices such as use of standardized/validated measures as discussed in the introduction section of the article. Articles were considered best practices if

1. They used standardized and well-validated measures as opposed to those that were ad hoc, had no prior validity information, or were known to be controversial (e.g., unstandardized versions of the Competitive Reaction Time Test). Measures were considered standardized if they had a clear protocol that must be followed by all scholars without deviation. Standardized tests reduce the potential for researcher degrees of freedom that create false-positive results (Przybylski & Weinstein, 2019b). Validated measures are those that have been demonstrated to predict outcomes related to clinically significant aggression (e.g., Child Behavior Checklist, Conflict and Tactics Scale) or criminal-justice outcomes (e.g., National Youth Survey, Negative Life Effects scales for delinquency). For experimental studies, demonstration of standardization and reliability resulted in coding as best practices given that few clinical validation studies have been attempted with laboratory measures of aggression. For correlational and longitudinal studies, in addition to demonstrating standardization and reliability, measures with validity coefficients of at least .20 in predicting outcomes such as clinical diagnoses, arrests, or other criminological outcomes were coded as best practices.
2. They controlled for theoretically relevant third variables (e.g., gender, trait aggression, mental health, family environment) in correlational and longitudinal studies. Studies were coded as best practices if they controlled for, at minimum, gender and at least one other relevant variable as well as Time 1 aggression in longitudinal studies.
3. They matched violent and nonviolent games carefully in experimental studies and avoided

using nonviolent games as violent-game exemplars. Games were considered to be matched carefully if the games were modified versions of the same game or provided evidence the games had been matched on relevant qualities such as difficulty, competitiveness, presence or absence of storyline, and difficulty learning controls and were reasonably close in year of release. Demonstration of matching could be achieved in multiple ways, either by consulting with an expert in such matters (e.g., Adachi & Willoughby, 2011b); using pretesting of game; using controls for difficulty, pace of action, and competitiveness in statistical analyses; or using "modding" of games or variants of the same game with violent and nonviolent scenarios.

We do acknowledge that what constitutes "best practices" can be subjective. Different scholars may adopt different views of what constitutes best practices and do so in good faith. For instance, other perspectives on best practices (e.g., C. A. Anderson, 2004) did not consider the issue of standardization, which we considered essential. Thus, we readily acknowledge that no best-practices approach necessarily can cover every issue or is beyond debate; however, we do believe they can function as a means of identifying whether a specific set of criteria could influence effect sizes in a field of research.

Finally, **[AQ: 8]**we also coded for the presence or absence of citation bias (the tendency to cite only studies supportive of the authors' hypotheses and ignore disconfirmatory studies). Studies were credited as avoiding citation bias if they cited a single study opposing the study's hypotheses, a fairly low bar. Citation bias can be one indicator of the impact of researcher-expectancy effects and has been found to correlate with higher effect sizes in other meta-analyses (Ferguson, 2015a). Although we had planned to include separate gender analyses, too few studies provided separate data for males and females.

Continuous moderator variables (age, date) were examined using metaregression. This technique allows for examination of a correlation between a continuous moderator and study effect size using regression techniques. Categorical moderators can be examined for subgroup differences in effect size that are significant (i.e., unlikely to be due to chance). This can be done with their fixed-effects or mixed-effects models. With mixed-effects models, as with random-effects models for overall meta-analysis, equal variance among studies is not assumed across subgroups. Thus, mixed-effects models in fields with heterogeneous study methods tend to be more appropriate, although we note both fixed- and mixed-effects models in the results section.

When differences occurred, mixed-effects models were preferred to fixed, although in the end, no differences emerged between models.

Interpretation of effect sizes has been controversial in violent-video-game research. Many such effect sizes are near zero but may be "statistically significant" because of the high power of meta-analyses. This may result in miscommunication as trivial effects become statistically significant (Orben & Przybylski, 2019). Although any cutoff is arbitrary, we set a cutoff of $r = .10$, below which the risk of effect sizes being explained primarily as being due to study artifacts rather than real population-level effects is quite high (Przybylski & Weinstein, 2019a).

Obviously, any strict cutoff is arbitrary. An effect (r) of .11 is not much different from one of .09. But we are concerned that meta-analyses too often rely on statistical significance as an indicator of hypothesis support despite the fact that the power of meta-analyses results in almost all outcomes being statistically significant. Recent scholarship has indicated that effects below .10 have a greater than 5% tendency to achieve statistical significance, particularly in large data sets despite the obvious "nonsense" nature of the variables involved such as the effect of potatoes on suicide (Orben & Przybylski, 2019), or the age one moved to a new city and tendency to play violent games ($r = −.14, p < .001$), or even hostility and myopia ($r = .04, p = .047$; Ferguson & Wang, 2019). Although more data would be welcome, it is likely that certain methodological issues, such as demand characteristics, single-responder bias, common methods variance, issues with survey-questionnaire wording, mischievous responding, and so on, can create a fairly systematic level of small false-positive results that become statistically significant in high-power studies, including meta-analyses. Thus, on the basis of advice of other scholars in the field (Orben & Przybylski, 2019; Przybylski & Weinstein, 2019a) and with an awareness of the arbitrary nature of any hard cutoff, we do not consider effect sizes below .10 to be evidentiary. Related to the arbitrary nature of this cutoff, our concern is less that such effect sizes may be evidentiary and more that effect sizes above .10 may still not be evidentiary. However, our observation is that the nonsense potential of an effect size diminishes significantly above .10 (but does not go away entirely).

Analysis

The Comprehensive Meta-Analysis software program was used to fit random-effects models. The potential for publication bias was assessed using the tandem procedure (Ferguson & Brannick, 2012), which looks for concordance among several funnel-plot-related tests for bias (Orwin's fail-safe N, Egger's regression, trim

Table 1. Meta-Analytic Results: Effect of Violent-Video-Game Exposure Studies on Outcome Variables

Variable	k	r_+	95% CI for r_+	Homogeneity test	I^2	τ	Publication bias?
Aggressive behavior	40	.09	[.07, .12]	$\chi^2(39) = 88.50, p < .001$	55.9	.06	No
Prosocial behavior	14	−.01	[−.07, .05]	$\chi^2(13) = 50.37, p < .001$	74.2	.096	No
Aggressive affect	12	.13	[.08, .18]	$\chi^2(11) = 10.60, p = .48$	00.0	.000	No
Aggressive cognitions	10	.13	[.03, .23]	$\chi^2(9) = 24.06, p = .004$	62.6	.127	No
Desensitization	7	.31	[.14, .46]	$\chi^2(6) = 19.40, p = .004$	69.1	.198	No

Note: k = number of studies; r_+ = pooled effect-size estimate; CI = confidence interval; I^2 = heterogeneity statistic; publication bias = decision based on the tandem procedure. The effect size r_c indicates the effect size corrected for publication bias as applicable.

and fill). This procedure is an empirically demonstrated, conservative estimating procedure for assessing publication bias, with low Type I error rates. However, it should be noted that by reducing Type I error rates, Type II error rates for the tandem procedure are higher. Thus, it should be considered a very specific but less sensitive measure for detecting publication bias. A negative result on the tandem procedure does not ensure the absence of publication bias. Assessments of publication bias were used on the basis of concordance of Orwin's fail-safe N (how many studies it would take to reduce effect sizes to $r = .10$, indicating fragility in the evidence base), Egger's regression for effect size and sample size, and the trim-and-fill procedure. Trim-and-fill corrections for publication bias, when warranted on the basis of the tandem procedure decision, are reported as r_c. The traditional fail-safe N, by focusing on statistical significance, typically vastly overestimates confidence in meta-analyses, but Orwin's version improves on this through an examination of effect sizes rather than statistical significance. Trim and fill, like most methods, typically has low power and the potential for Type II error (Ferguson & Brannick, 2012).

Publication bias was also assessed using *p*-curve analysis (Simonsohn, Nelson, & Simmons, 2014), looking for an overabundance of *p* values near to the $p = .05$ threshold. A clustering of effects near .05 can indicate *p*-hacking or massaging of data to cross the threshold into statistical significance, particularly among smaller experimental studies. Last, the R index (Schimmack, 2014), or the degree to which significant results match the frequency expected given the power of the studies, was also used. Such an index can reveal whether a pattern of significant results in studies is unlikely given the observed power of those studies. The R index provides an estimate of the strength of evidence for a hypothesis and whether that strength may have been exaggerated through questionable researcher practices. Lower R index scores generally suggest that replicability of hypothesis-supportive findings is less likely, either because observed findings exceed observed power or observed findings were low

in initial support. Although newer, the R index has seen increasing use as part of meta-analysis (e.g., Harms, Genau, Meschede, & Beauducel, 2018; Lynn, 2018; Sakaluk, Williams, Kilshaw, & Rhyner, 2019; for discussion, see also De Boeck & Jeon, 2018).

Results

Table 1 presents the overall results for all study types across the various outcome realms. Publication bias results here are based on the tandem procedure only. We report further publication bias analyses in a later section. Results suggested minimal effects for aggressive behavior, no effects at all for prosocial behavior, very small effects for aggressive affect and cognition, but comparatively large effects for physiological desensitization.

Given that studies of aggressive behavior are both most numerous and, arguably, most relevant to the core debate, the following analyses focus on these specifically. Table 2 examines effect sizes for studies of aggressive behavior across study types. As can be seen, the strongest evidence for effects on aggressive behavior comes from experimental studies; the weakest evidence is from longitudinal studies. Experimental study effects are small, whereas effects for correlational and longitudinal studies are below the level of interpretation. Although the differences among study types were significant in a fixed-effects analyses, $Q(2) = 7.49, p = .024$, results were nonsignificant in mixed-effects analyses.[2]

Categorical moderators are considered in Table 3. Effects were higher both for studies with NBPs and for studies with citation-bias issues. Both fixed- and mixed-effects-model analyses were significant for best practices—mixed-effects data reported were $Q(1, 38) = 14.36, p < .001$. Likewise, both fixed- and mixed-effects analyses were significant for citation bias—mixed-effects data reported were $Q(1, 38) = 12.86, p < .001$. These results indicate that the moderators were significant predictors for effect size. Both of these groups of studies also demonstrated evidence of publication bias. Effect sizes were weaker in best-practices studies and studies with more balanced literature reviews. Effects from

Table 2. Meta-Analytic Results: Effect of Violent-Video-Game Use on Aggressive Behavior Across Study Types

Variable	k	r_+	95% CI for r_+	Homogeneity test	I^2	τ	Publication bias?
Correlational	14	.08	[.04, .12]	$\chi^2(13) = 31.33, p = .003$	58.5	.051	No
Longitudinal	8	.07	[.04, .11]	$\chi^2(7) = 9.67, p = .208$	63.7	.027	No
Experimental	18	.14	[.07, .21]	$\chi^2(17) = 40.01, p < .001$	57.5	.115	No

Note: k = number of studies; r_+ = pooled effect-size estimate; CI = confidence interval; I^2 = heterogeneity statistic; publication bias = decision based on the tandem procedure. The effect size r_c indicates the effect size corrected for publication bias as applicable.

best-practices experimental studies specifically were nonsignificant ($k = 5$, $r = -.025$; $Q = 3.38$, $p = .496$). This suggests that best-practices studies do not provide evidence in support of the hypothesis that violent game playing is associated with aggression. Neither the age moderator, $Q(1, 34) = 0.25$, $p = .62$, nor the study year moderator, $Q (1, 38) = 0.37$, $p = .54$, proved significant in metaregression analyses.

Publication-bias analyses

As indicated in the earlier analyses, the tandem procedure revealed evidence for publication bias among NBP experimental studies of aggressive behavior as well as studies with citation bias. We also sought to examine for potential issues using *p*-curve and R index analyses. However, these are most appropriate for pools of studies with high rates of statistical significance. Given a plethora of null findings in the current pool of data, their use did not appear to be warranted. We can provide outcome data for scholars who are interested in these analyses on request. Both tools generally are appropriate for detecting when the number of statistically significant findings is higher than expected given the observed power of the included studies. However, when there are many null results or very weak effect

sizes, neither the *p*-curve nor R index is likely to return meaningful results.

Finally, in Table 3, we document the difference in effect-size estimates across several meta-analytic variations from the pool of studies available to the 2015 APA task force. The first column indicates the results reported by the APA task force in its technical report (APA, 2015b). The second column represents a direct replication of only the studies from the APA task-force meta-analysis for which effect sizes could be calculated (e.g., minus the five studies for which no effect size relevant to hypotheses about the effects of violent games on aggression could be calculated). The third column represents the expanded reanalysis with a wider range of studies included. The fourth column represents the effect sizes for the best-practices studies included in the current reanalysis.

This comparison involves categories used by both the APA 2015 task force and the current meta-analysis. The 2015 task force chose to collapse aggressive affect with prosocial behaviors, so our comparison does the same. As can be seen, even a direct replication produces weaker effect sizes than reported in the 2015 task-force report (APA, 2015b). Best-practices studies have the weakest effect sizes of all, with results that were not statistically different from zero. Only in the case of

Table 3. Moderator Analysis for Categorical Moderators of Studies of Aggressive Behavior

Variable	k	r_+	r_c	95% CI for r_+	Homogeneity test	I^2	τ	Publication bias?
Citation bias								
Yes	19	.14	.07	[.11, .18]	$\chi^2(18) = 30.09, p = .037$	40.2	.049	Yes
No	21	.05		[.02, .09]	$\chi^2(20) = 42.01, p = .003$	52.4	.053	No
Best practices								
Yes	14	.04		[.01, .07]	$\chi^2(13) = 14.28, p = .35$	8.9	.017	No
No	26	.13	.07	[.09, .16]	$\chi^2(24) = 54.59, p < .001$	54.2	.058	Yes
Best-practices effects								
Aggressive behavior	14	.04		[.01, .07]	$\chi^2(13) = 14.28, p = .35$	8.9	.017	No
Prosocial behavior	6	−.02		[−.12, .09]	$\chi^2(5) = 8.29, p = .14$	39.7	.08	No
Aggressive affect	3	.08		[−.02, .18]	$\chi^2(2) = 1.85, p = .40$	0.0	.0	No

Note: k = number of studies; r_+ = pooled effect-size estimate; r_c = effect size corrected for publication bias; CI = confidence interval; I^2 = heterogeneity statistic; publication bias = decision based on the tandem procedure.

Table 4. Meta-Analytic Results Across Different Data Sets (2009–2013)

Variable	Original task force	Directly replicable studies	Reanalysis	Best-practices reanalysis
Aggressive behavior	.184	.097	.093	.038
Aggressive affect	.093	.092*	.051*	.019*
Aggressive cognitions	.166	—	.129	—
Desensitization	.189	—	.307	—

Note: The 2015 task force appeared to collapse aggressive affect and empathy/prosocial into a single category. For this comparison, we similarly aggregated effect sizes for affect and prosocial outcomes. The task force's category for physiological arousal and our use of desensitization also appeared similar, and the term *desensitization* is maintained for this category, although this was described as physiological arousal in the original task force report. Only a single study of cognition as outcome was considered best practices, making comparison difficult. — = too few studies for meta-analysis. *$p < .05$. [AQ: 9]

desensitization did effect sizes increase in the current reanalysis, but this appeared driven by NBP studies.

Exploratory analyses

An exploratory analysis was conducted to examine whether there were any differences among the studies included in the 2015 task-force meta-analysis, rejected from the meta-analysis, and discovered with the new broadened search. Outcomes were related to study type (experimental, correlational, and longitudinal) and whether studies were rated as best practices or NBPs. The analysis was conducted at the level of studies rather than articles.

Results indicated that the three inclusion groups (included, rejected, and new) did not differ in respect to the types of studies included, $\chi^2(4) = .859$, $p = .930$. However, it is worth noting that overall, experiments were overrepresented (65.8% of studies vs. 23.3% for correlational and 11.0% for longitudinal). For best practices, results across inclusion groups were significant, $\chi^2(7) = 7.541$, $p = .023$. Specifically, the percentage of best practices studies was much higher in the rejected group (45%) than in either the group of studies originally included in the APA task-force meta-analysis (22.2%) or studies found in a new search of the literature (12.5%).

As noted earlier, meta-analyses that include only bivariate correlations risk overestimating the evidence for a hypothesis. It would be possible for a series of studies employing multivariate analyses with controls to find null results in every study, yet a meta-analysis of bivariate correlations from those studies to conclude the hypothesis is supported. Thus, meta-analyses of better controlled effect sizes are often superior to bivariate correlations, particularly given evidence that assumptions of the superiority of bivariate correlations for meta-analysis appear to be incorrect (Furuya-Kanamori & Doi, 2016). Nonetheless, as a point of comparison, we include effect sizes for the original task-force report (APA, 2015b), bivariate correlations from our analyses, and the controlled effects from our analyses. Given that this difference between controlled and bivariate effects is less relevant for experimental studies, we limited our comparison to correlational and longitudinal studies. Most such studies were also in the realm of aggressive behavior. There were too few studies in other realms for meaningful meta-analysis.

Results are included as Table 5. As can be seen, our results for bivariate analyses largely replicated those of the original APA task-force report (APA, 2015b) despite the latter having included only three studies of aggressive behavior that were correlational or longitudinal. In bivariate analyses, about 2.5% of the variance in aggression is explained by violent video games. In controlled analyses, that shrinks considerably to approximately 0.5% of the variance and is a level some scholars caution against interpreting as hypothesis supportive (e.g., Orben &

Table 5. Comparison of Effect Sizes Across Bivariate and Better-Controlled Analyses

Original task force report	Current study bivariate	Current study controlled
$r = .16$ ($k = 3$)	Correlational: $r = .15$ ($k = 13$)	Correlational: $r = .08$ ($k = 14$)
	Longitudinal $r = .16$ ($k = 8$)	Correlational: $r = .07$ ($k = 8$)

Note: Because the original American Psychological Association task-force report contained only three studies, longitudinal and correlational effect sizes were combined.

Przybylski, 2019) because of the possibility that such effect sizes are likely to be artifacts, not true effects. This demonstrates the importance of understanding controlled effect sizes regarding a hypothesis and how bivariate effect sizes can provide overconfidence in regard to whether a hypothesis has been supported in the evidence base.

Discussion

In 2015, the APA task force on violent video games released its technical report, which was based on a meta-analysis of 18 violent-video-game studies (APA, 2015b). This technical report became the foundation for a resolution by the APA concluding that such games were conclusively linked to aggressive behavior, although not violent behavior. In the current reanalysis, we sought to replicate the original meta-analysis that formed the basis for these conclusions. It was our intent to examine the extant literature in 2015 that informed the APA's resolution on violent games to examine whether the APA's resolution statement accurately portrayed the field at that time. On balance, we found the original meta-analysis to be difficult to replicate. Five of the original 18 studies (28% of the studies used in the task force's meta-analysis) appeared not to have sufficient data from which to extract meaningful effect sizes relevant to the hypothesis that violent games are linked to aggression. By contrast, we included a large number of studies that had been missed by the original task force, bringing the total pool of studies for the relevant time frame to 62 studies (not including the five we excluded from the original APA task-force analysis for insufficient data, as noted above).

Our reexamination found weaker overall evidence for effects of violent video games on aggression, with the exception of physiological desensitization. One weakness of meta-analysis is that such procedures may inadvertently mask significant heterogeneity in between-study results, leading to overconfidence in the consistency of results, and we are concerned this is the case for this pool of data. Experimental studies of aggressive behavior with NBPs and studies that engaged in citation bias tended to have higher effect sizes than higher-quality studies. On balance, the evidence from this reanalysis suggests multiple points of evidence that a more cautious set of conclusions regarding the impact of violent games on aggression would have been warranted than were evidenced in the 2015 task-force report (APA, 2015b) and subsequent APA resolution statement (APA, 2015a).

We do wish to communicate our appreciation for the 2015 task-force members' efforts in attempting to synthesize a difficult and contentious body of literature. Although we are, at times, critical of their efforts, we recognize the inherent difficulty of meta-analyses. Many decisions made in such analyses can take the form of good-faith differences in approach. Furthermore, we recognize that many of the potential problems we identify for the 2015 task force are likely common among meta-analysis and, in some cases (e.g., the use of bivariate rs), have even become established practice, although potentially resulting in faulty conclusions. Thus, we view this effort not so much as intending to be critical of some individual scholars but as a case study in how the often regular practice of meta-analysis can lead to faulty conclusions.

Many of the issues that appear to influence effect sizes in this field have been known for at least a decade. These include the unstandardized use of aggression measures, poor matching of experimental conditions (Adachi & Willoughby, 2011b), poor controls in correlational studies, and overreliance on upwardly biased bivariate correlations in meta-analyses (Furuya-Kanamori & Doi, 2016). The open letter written to the APA and the 2015 task force by 230 scholars (Consortium of Scholars, 2013) noted many of these concerns. Our own reanalysis suggests these issues are not trivial but rather, are associated with higher effect sizes that may spuriously inflate confidence in effects. It is possible that greater attention to these issues may have revised the task force's conclusions.

For the central question of the impact of violent video games on aggressive behavior, the effect size (r) for all studies was .09, with larger effect sizes for experimental studies than correlational and longitudinal studies. Examining these experimental studies more closely, they appear to be experiencing publication bias, particularly among studies with NBPs or citation bias. Adjusted effect sizes for better-quality studies or for publication bias suggest small effects in the range of .01 to .07. This would bring experimental studies more in line with the correlational and longitudinal studies, which suggest overall effects that are generally close to zero. This means that the variance in violent game play and aggressive behavior overlaps by less than half a percent. This is in line with other effects that are generally regarded as trivial and unimportant (Orben & Przybylski, 2019) and could be noise resulting from methodological artifacts rather than real effects.

One issue is whether the greater number of correlational studies in our analysis relative to the original APA task-force meta-analysis may explain some of the difference in our results. As indicated above, it does appear that experimental studies do demonstrate higher effect sizes (although at $r = .14$, our results are still lower than for the original task-force report, APA, 2015b) than correlational ($r = .08$) or longitudinal ($r = .07$) designs, which suggests that giving higher weight to experimental studies tends to increase effect sizes. However, it is not possible to conclude that experiments

necessarily provide better evidence for causal effects compared with other studies. Pools of NBP experimental studies demonstrate publication bias, and adjusting for this tends to pull observed effect sizes down closer to those seen for correlational and longitudinal studies. The effect size for best-practices experimental studies ($k = 5$, $r = -.025$; $Q = 3.38$, $p = .496$) is actually *negative*, which suggests that most of the positive results in experimental studies can be explained via either publication bias or NBPs. This is consistent with other prior analyses (Hilgard, Engelhardt, & Rouder, 2017). Thus, we conclude it is less that the inclusion of correlational and longitudinal studies is at issue but rather that the field has tended to give undo weight to poor-quality studies without considering quality issues or the potential for publication bias.

Very small effects were found for aggressive cognition or affect. There were too few such studies to conduct an in-depth analysis of statistical methods that might bias effect sizes, such as demand characteristics. Thus, these effects should be interpreted with some caution.

For prosocial behaviors and physiological desensitization, opposing results were found. Results for prosocial behaviors were effectively no different from zero. By contrast, studies of physiological desensitization were far stronger than any other effects. It is interesting to consider why one outcome stands out compared with others. These results come from only six effect sizes across four articles, none of which were rated as best practices, so it is possible that this effect size is not representative of true population effects. It is also possible that the outcome measures in this field may be tapping into an entirely different construct from aggressiveness. We feel that this is worth examining more carefully, particularly using standardized measures and with preregistered open science designs.

The observation that the 2015 task-force meta-analysis rejected a high proportion of best-practices studies is also a matter of concern. Indeed, best-practices studies were more common among studies the task force rejected than among studies it included. Ultimately, few meta-analyses are likely perfect in reliably making inclusion and exclusion decisions, so we do not intend to be overly critical here. However, it is possible that the inclusion of more best-practices studies may have had an impact on the task force's conclusions.

Moving forward

Ostensibly, by convening a task force and engaging meta-analysis, the APA sought to resolve controversies within the field as well as to elucidate a clear position to resolve debate in the public. Unfortunately, in this, the APA was not successful and, arguably, mainly succeeded in becoming embroiled in further controversy regarding the composition and potential biases of the task force and the degree to which the APA's resolution could be supported by data available at that time. None of this should be taken to imply bad faith, but rather that by seeking to provide "the answer" to a contentious and heterogeneous field, the APA's efforts may have been of limited pragmatic value and, instead, created the optics of it *taking sides* in a manner that was not necessarily well informed by data. The current reanalysis suggests that on the whole, contemporarily available data would have argued for a more cautious set of conclusions than were promoted by the APA task force (APA, 2015b) and subsequent APA resolution (APA, 2015a). This has several ramifications for the field moving forward.

Since the APA task-force technical report was released, several other meta-analyses have been released, each coming to different conclusions (e.g., Ferguson, 2015a; Furuya-Kanamori & Doi, 2016; Hilgard et al., 2017; Prescott et al., 2018). Most of these had rather specific goals. For instance, Ferguson (2015a) examined only studies of youths and children. Furuya-Kanamori and Doi (2016) replicated Ferguson but was also specifically interested in examining the applicability of standardized regression coefficients. Hilgard et al. (2017) was a reanalysis of an older meta-analysis (C. A. Anderson et al., 2010), and the authors discovered that results from experimental studies in this older meta-analysis were mainly the product of publication bias. Prescott et al. (2018) concerned itself with longitudinal studies only and, like this analysis, used standardized regression coefficients. None of these other meta-analyses were involved with reexamining the APA 2015 task-force results. A recent summary of meta-analyses (Mathur & VanderWeele, 2019; for commentary, see also Drummond & Sauer, 2019) found that effect sizes were uniformly quite small, under $r = .10$, with systematic methodological issues as likely to explain outcomes as real effects. Scholars have cautioned against interpreting such small effects as hypothesis supporting given the high potential for false positives as a result of methodological issues within that effect-size range (Przybylski & Weinstein, 2019b). Finally, our best-practices analysis confirms that the quite small effects found by Mathur and VanderWeele (2019) were likely driven by methodological problems given that best-practices studies' effects were little different from zero.

Meta-analysis. The results from the current study remind one that the interpretation of meta-analyses needs to be nuanced. The pooled mean effect sizes are not always an accurate representation of population-level effects, particularly when effect sizes are driven higher by NBPs. Likewise, it is possible to achieve statistically

significant nonzero effect sizes, even when a majority of studies fail to provide evidence in favor of a hypothesis.

In the future, meta-analysts ought to avoid overreliance on mean effect sizes in which methodological issues are known to be prevalent. Likewise, we advise against the use of bivariate correlations in correlational and longitudinal samples because these also can be upwardly biased. Meta-analysis can be an effective tool in examining study-level factors that can drive higher or lower effect sizes. However, we believe it is erroneous to assume that the pooled mean effect size could be interpreted meaningfully and as supportive of effects. Put bluntly, meta-analytic authors should desist in concluding that *the average effect size wins*.

Perhaps most critically, both meta-analysis and individual studies, particularly with large samples, should stop focusing on statistical significance as a benchmark for hypothesis support. For over two decades, overreliance on statistical significance has been known to artificially increase confidence in weak results (Wilkinson & Task Force on Statistical Inference, 1999). Furthermore, we can now see that the effect sizes produced by some meta-analyses in this field, including not only this data set, are no different in magnitude from nonsense correlations (see Ferguson & Wang, 2019), such as the relationship between age one has moved to a new city and aggression or between potato consumption or wearing eyeglasses and mental health (Orben & Przybylski, 2019). Thus, it is clear that overinterpretation of extremely small effect sizes, particularly below .10, is likely to reify many false-positive results even if statistically significant. In some cases, this may be due to traditional Type I error, sampling issues, or methodological issues (e.g., demand characteristics, common methods variance, etc.). In accordance with recommendations by Orben and Przybylski (2019), at least within media psychology, extremely small effect sizes below this magnitude should no longer be interpreted as hypothesis supportive even if statistically significant.

Open science. The APA task-force meta-analysis relied on a pool of studies produced before open science principles such as preregistration had become common. Quality issues, including questionable researcher practices, appear to be common in this pool of studies. To be fair, such issues were widespread across many fields in psychology. However, the difference in effect sizes between best-practices studies (which were more standardized) and NBP studies suggests effect sizes may have been inflated by quality issues in many studies. Since this time, a limited number of preregistered studies have been produced (e.g., Ferguson, Trigani, et al., 2015; A. H. Ivory, Ivory, & Lanier, 2017; McCarthy, Coley, Wagner, Zengel, & Basham, 2016; Przybylski & Weinstein, 2019a). More are certainly desirable. It may be possible in future years to examine differences in outcome between preregistered and nonpreregistered studies. Currently, the pattern suggests that preregistered studies are much less likely to find evidence for adverse effects (Przybylski & Weinstein, 2019b), but a higher volume of studies is certainly welcome.

With that in mind, we encourage future researchers to take advantage of preregistration. Such a process can help cut back on researcher-expectancy effects that, as evidenced by citation bias, appear to influence effect sizes in the field. We are aware that poor-quality preregistrations such as those with few details may not be sufficient to reduce researcher-expectancy effects, but we argue that preregistration is definitely a step in the right direction toward greater clarity for the field.

Limitations

As with any study, this one has limitations. For meta-analyses, the quality of the results is only equal to the quality of studies included. Although we engaged in some analyses of practices that could influence effect sizes, ultimately the GIGO problem is unavoidable, particularly for a field well known for significant methodological controversies. Finally, although the use of standardized regression coefficients is in many ways superior to the use of bivariate correlations, there is a potential for model misspecification in some regression models. In such cases, model misspecification might spuriously inflate effect-size estimates. This can happen when "control" variables bear little theoretical relevance to the outcome or predictor variables, creating an illusion of control, or through multicollinearity, which can create inflation in the size of regression coefficients.

One other note is worth considering. Regarding best practices, it is quite likely that many studies claim best practices despite engaging in problematic practices (unstandardized measures, poor matching of games, poor statistical controls, obvious demand characteristics, etc.). The epitome perhaps is the "best practices" analysis from an older meta-analysis (C. A. Anderson et al., 2010), which produced a pool of studies that commonly used unstandardized measures, poorly matched games, poor controls, and that ultimately were found to largely be the product of publication bias (Hilgard et al., 2017). Thus, scholars should become alert for specific practices that are rigorous rather than relying on authors' claims of engaging in best practices.

Concluding thoughts

In 2020, the APA clarified its policy statement on video games, adding a statement that the resolution should not be used to link video games to violent behavior. We applaud this clarifying comment, which we believe will reduce the degree to which the APA's statement

has been used to distract from other issues such as gun control. However, the APA retained its position that there is "consistent" evidence linking games to aggression (left undefined) despite significant heterogeneity and controversy in the evidence base. In a rare rebuke, the APA's own division for media research, the Society for Media and Technology, released an open letter critical of the APA's resolution, asking the APA to retire it (Society for Media Psychology and Technology, 2020). This illustrates the degree to which the APA's resolution on games has done more to open it to criticism rather than resolve ongoing debates.

Our current analysis suggests that a careful examination of the evidence available at the time of the 2015 APA task-force technical report would have warranted considerable caution in linking violent video games to aggressive behavior. Unfortunately, the current APA resolution derived from the technical report likely misrepresents the research field considerably and greatly overestimates the degree of consistency among studies in this field, most of which in fact are incapable of supporting links between violent games and aggression. Again, our intent is not to imply any bad faith on the part of the 2015 task-force members, and we appreciate the hard work they put into synthesizing a difficult field. However, we express the concern that the ultimate result perhaps did more harm than good to the field and that the resolution stemming from it continues to misinform the general public, although certainly unintentionally. We recommend that the APA reconsider its position related to video-game violence and retire and archive the 2015 resolution on video-game violence.
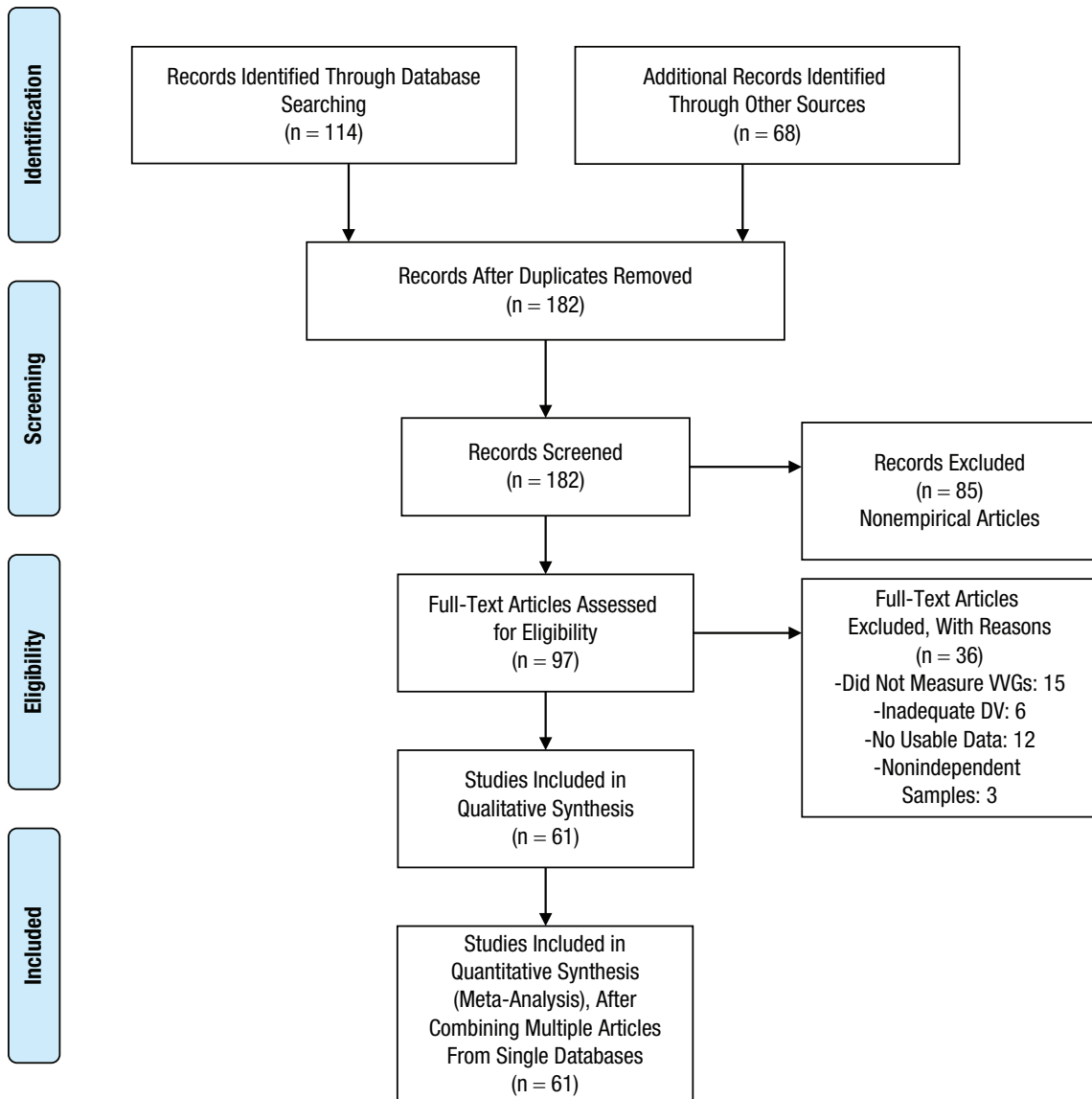
Appendix

Fig. A1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) chart.

Transparency

Action Editor: Richard Lucas
Editor: Laura A. King
Author Contributions

C. J. Ferguson conceived of the original study and was involved in manuscript selection and coding and data analysis. A. Copenhaver was involved in conceiving the original design and study selection. P. Markey was involved in effect-size coding and data analysis. All of the authors contributed to writing the manuscript and approved the final version for submission.

Declaration of Conflicting Interests

The author(s) declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

ORCID iD

Christopher J. Ferguson [iD] https://orcid.org/0000-0003-0986-7519

Acknowledgments

The analysis plan outline can be found at https://osf.io/7cjn6/ with a permanent copy at https://osf.io/g49af. Data from the meta-analysis can be found at https://osf.io/fpua4/. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist can be found at https://osf.io/x92pb/.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/1745691620927666

Notes

1. One exception was two partially overlapping longitudinal samples (Ferguson, 2011; Ferguson, San Miguel, Garza, & Jerabeck, 2012) drawn from a larger correlational database. Analyses were run with both samples and the latter 2012 sample only. Results did not significantly differ either way. Reported results in this article are for both subsamples.
2. Note that these refer to computation options for examining group differences as opposed to the overall meta-analytic model.

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*, 259–274.

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*, 55–62. doi:10.1016/j.avb.2010.12.002

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

American Psychological Association. (2015a). APA review confirms link between playing violent video games and aggression. Retrieved from https://www.apa.org/news/press/releases/2015/08/violent-video-games.aspx

American Psychological Association. (2015b). APA Task Force on Violent Media: Technical report on the review of the violent video game literature. Retrieved from https://www.apa.org/pi/families/review-video-games.pdf

Anderson, C. A. (2004). An update on the effects of playing violent video games. *Journal of Adolescence*, *27*, 113–122.

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

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.

Anderson, M. (2018). Trump reopens a seemingly settled debate on video games and violence. Retrieved from https://www.pbs.org/newshour/politics/trump-reopens-a-seemingly-settled-debate-on-video-games-and-violence

Beerthuizen, M. G. C. J., Weijters, G., & van der Laan, A. M. (2017). The release of grand theft auto V and registered juvenile crime in the Netherlands. *European Journal of Criminology*, *14*, 751–765. doi:10.1177/1477370817717070

Bowman, N. (2012). Effect sizes and statistical methods for meta-analysis in higher education. *Research in Higher Education*, *53*, 375–382.

Breuer, J., Vogelgesang, J., Quandt, T., & Festl, R. (2015). Violent video games and physical aggression: Evidence for a selection effect among adolescents. *Psychology of Popular Media Culture*, *4*, 305–328. doi:10.1037/ppm0000035

Brown v. Entertainment Merchants Association, 131 S. Ct. 2729. (2011). Retrieved from http://www.supremecourt.gov/opinions/10pdf/08-1448.pdf

Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin*, *28*, 1679–1686.

Calvert, S. L., Appelbaum, M., Dodge, K. A., Graham, S., Nagayama Hall, G. C., Hamby, S., . . . Hedges, L. V. (2017). The American Psychological Association Task Force assessment of violent video games: Science in the service of the public interest. *American Psychologist*, *72*, 126–143.

Consortium of Scholars. (2013). *Scholar's open statement to the APA task force on violent media*. Retrieved from http://www.scribd.com/doc/223284732/Scholar-s-Open-Letter-to-the-APA-Task-Force-On-Violent-Media-Opposing-APA-Policy-Statements-on-Violent-Media

Copenhaver, A. (2015). Violent video game legislation as pseudo-agenda. *Criminal Justice Studies: A Critical Journal of Crime, Law and Society*, *28*, 170–185. doi:10.1080/1478601X.2014.966474

Copenhaver, A., & Ferguson, C. J. (2018). Selling violent video game solutions: A look inside the APA's internal notes leading to the creation of the APA's 2005 resolution on violence in video games and interactive media. *International Journal of Law and Psychiatry*, *57*, 77–84.

Crossley, R. (2014). Mortal Kombat: Violent game that changed video games industry. Retrieved from https://www.bbc.com/news/technology-27620071

Cunningham, S., Engelstätter, B., & Ward, M. R. (2016). Violent video games and violent crime. *Southern Economic Journal*, *82*, 1247–1265. doi:10.1002/soej.12139

De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. *Psychological Bulletin*, *144*, 757–777. doi:10.1037/bul0000154

DeCamp, W. (2015). Impersonal agencies of communication: Comparing the effects of video games and other risk factors on violence. *Psychology of Popular Media Culture*, *4*, 296–304. doi:10.1037/ppm0000037

DeCamp, W., & Ferguson, C. J. (2017). The impact of degree of exposure to violent video games, family background, and other factors on youth violence. *Journal of Youth and Adolescence*, *46*, 388–400.

DeLisi, M., Vaughn, M. G., Gentile, D. A., Anderson, C. A., & Shook, J. J. (2013). Violent video games, delinquency, and youth violence: New evidence. *Youth Violence and Juvenile Justice*, *11*, 132–142. doi:10.1177/1541204012460874

Del Re, A., & Hoyt, W. (2019). Aggregate dependent effect sizes. Retrieved from https://rdrr.io/cran/MAd/man/agg.html

Devilly, G. J., Brown, K., Pickert, I., & O'Donohue, R. (2017). An evolutionary perspective on cooperative behavior in gamers. *Psychology of Popular Media Culture*, *6*, 208–221. doi:10.1037/ppm0000097

Drummond, A., & Sauer, J. (2019). Divergent meta-analyses do not present uniform evidence that violent video game content increases aggressive behaviour. *PsyArXiv*. Retrieved from https://psyarxiv.com/xms5u/

Ducharme, J. (2018, March 12). Trump blames video games for school shootings. Here's what science says. TIME Magazine. Retrieved from http://time.com/5191182/trump-video-games-violence/

Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (2015). Comparing apples and oranges? Evidence for pace of action as a confound in research on digital games and aggression. *Psychology of Popular Media Culture*, *4*, 112–125. doi:10.1037/ppm0000010

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*, *26*, 432–419. doi:10.1037/a0035569

Elson, M., & Quandt, T. (2016). Digital games in laboratory experiments: Controlling a complex stimulus through modding. *Psychology of Popular Media Culture*, *5*, 52–65. doi:10.1037/ppm0000033.supp

Ferguson, C. J. (2011). Video games and youth violence: A prospective analysis in adolescents. *Journal of Youth and Adolescence*, *40*, 377–391.

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*, 57–74.

Ferguson, C. J. (2015a). Do angry birds make for angry children? A meta-analysis of video game influences on children's and adolescents' aggression, mental health, prosocial behavior, and academic performance. *Perspectives on Psychological Science*, *10*, 646–666.

Ferguson, C. J. (2015b). Pay no attention to that data behind the curtain: On angry birds, happy children, scholarly squabbles, publication bias and why betas rule metas. *Perspectives on Psychological Science*, *10*, 683–691.

Ferguson, C. J., Barr, H., Figueroa, G., Foley, K., Gallimore, A., LaQuea, R., . . . Garza, A. (2015). Digital poison? Three studies examining the influence of violent video games on youth. *Computers in Human Behavior*, *50*, 399–410. doi:10.1016/j.chb.2015.04.021

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling and implications for the use of meta-analyses. *Psychological Methods*, *17*, 120–128.

Ferguson, C. J., San Miguel, C., Garza, A., & Jerabeck, J. (2012). A longitudinal test of video game violence effects on dating violence, aggression and bullying: A 3-year longitudinal study of adolescents. *Journal of Psychiatric Research*, *46*, 141–146.

Ferguson, C. J., Trigani, B., Pilato, S., Miller, S., Foley, K., & Barr, H. (2015). Violent videogames don't increase hostility in teens but they do stress girls out. *Psychiatric Quarterly*, *87*, 49–56.

Ferguson, C. J., & Wang, C. K. J. (2019). Aggressive video games are not a risk factor for future aggression in youth: A longitudinal study. *Journal of Youth and Adolescence*, *48*, 1439–1451.

Furuya-Kanamori, F., & Doi, S. (2016). Angry birds, angry children and angry meta-analysts. *Perspectives on Psychological Science*, *11*, 408–414.

Gerstenzang, J. (1999, April 25). Clinton sees violent influence in 3 video games. *Los Angeles Times*. Retrieved from http://articles.latimes.com/1999/apr/25/news/mn-30999

Greitemeyer, T., & Sagioglou, C. (2017). The longitudinal relationship between everyday sadism and the amount of violent video game play. *Personality and Individual Differences*, *104*, 238–242. doi:10.1016/j.paid.2016.08.021

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

Happ, C., Melzer, A., & Steffgen, G. (2013). Superman v. BAD man? The effects of empathy and game character in violent video games. *Cyberpsychology, Behavior, and Social Networking*, *16*, 774–778. doi:10.1089/cyber.2012.0695

Harms, C., Genau, H., Meschede, C., & Beauducel, A. (2018). Does it actually feel right? A replication attempt of the rounded price effect. *Royal Society Open Science*. Retrieved from https://royalsocietypublishing.org/doi/full/10.1098/rsos.171127

Hilgard, J., Engelhardt, C. R., & Rouder, J. N. (2017). Overstated evidence for short-term effects of violent games on affect and behavior: A reanalysis of Anderson

et al. (2010). *Psychological Bulletin*, *143*, 757–774. doi:10.1037/bul0000074

Hilgard, J., Engelhardt, C. R., Rouder, J. N., Segert, I. L., & Bartholow, B. D. (2019). Null effects of game violence, game difficulty, and 2D:4D digit ratio on aggressive behavior. *Psychological Science*, *30*, 606–616. doi:10.1177/0956797619829688

Ivory, A. H., Ivory, J. D., & Lanier, M. (2017). Video game use as risk exposure, protective incapacitation, or inconsequential activity among university students: Comparing approaches in a unique risk environment. *Journal of Media Psychology: Theories, Methods, and Applications*, *29*, 42–53. doi:10.1027/1864-1105/a000210

Ivory, J. D., Markey, P. M., Elson, M., Colwell, J., Ferguson, C. J., Griffiths, M. D., . . . Williams, K. D. (2015). Manufacturing consensus in a diverse field of scholarly opinions: A comment on Bushman, Gollwitzer, and Cruz (2015). *Psychology of Popular Media*, *4*, 222–229. doi:10.1037/ppm0000056

Kneer, J., Elson, M., & Knapp, F. (2016). Fight fire with rainbows: The effects of displayed violence, difficulty, and performance in digital games on affect, aggression, and physiological arousal. *Computers in Human Behavior*, *54*, 142–148. doi:10.1016/j.chb.2015.07.034

Krcmar, M., & Farrar, K. (2009). Retaliatory aggression and the effects of point of view and blood in violent video games. *Mass Communication & Society*, *12*, 115–138.

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

Kühn, S., Kugler, D., Schmalen, K., Weichenbergera, M., Witt, C., & Gallinat, J. (2018a). The myth of blunted gamers: No evidence for desensitization in empathy for pain after a violent video game intervention in a longitudinal fMRI study on non-gamers. *Neurosignals*, *26*, 22–30.

Kühn, S., Kugler, D. T., Schmalen, K., Weichenberger, M., Witt, C., & Gallinat, J. (2018b). Does playing violent video games cause aggression? A longitudinal intervention study. *Molecular Psychiatry*, *24*, 1220–1234. doi:10.1038/s41380-018-0031-7

Lin, J. H. (2013). 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*, 535–543.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Newbury Park, CA: Sage.

Lobel, A., Engels, R. C. M. E., Stone, L. L., Burk, W. J., & Granic, I. (2017). Video gaming and children's psychosocial wellbeing: A longitudinal study. *Journal of Youth and Adolescence*, *46*, 884–897. doi:10.1007/s10964-017-0646-z

Lynn, M. (2018). Are published techniques for increasing service-gratuities/tips effective? P-curving and R-indexing the evidence. *International Journal of Hospitality Management*, *69*, 65–74. doi:10.1016/j.ijhm.2017.10.004

Markey, P. M., Markey, C. N., & French, J. E. (2015). Violent video games and real-world violence: Rhetoric versus data. *Psychology of Popular Media Culture*, *4*, 277–295.

Mathur, M. B., & VanderWeele, T. J. (2019). Finding common ground in meta-analysis "wars" on violent video games. *Perspectives on Psychological Science*, *14*, 705–708. doi:10.1177/1745691619850104

McCarthy, R. J., Coley, S. L., Wagner, M. F., Zengel, B., & Basham, A. (2016). Does playing video games with violent content temporarily increase aggressive inclinations? A pre-registered experimental study. *Journal of Experimental Social Psychology*, *67*, 13–19. doi:10.1016/j.jesp.2015.10.009

McCarthy, R. J., & Elson, M. (2018). A conceptual review of lab-based aggression paradigms. *Collabra: Psychology*, *4*(1), 4. doi:10.1525/collabra.104

Olson, S. (2018, May 19). Guest opinion: Violent video games are teaching kids to kill. *Reporter-Herald*. Retrieved from http://www.reporterherald.com/opinion/guest-columns/ci_31889271/guest-opinion-violent-video-games-are-teaching-kids

Orben, A., & Przybylski, A. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behavior*, *3*, 173–182. doi:10.1038/s41562-018-0506-1

Pratt, T., Cullen, F., Sellers, C., Winfree, T., Madensen, T., Daigle, L., . . . Gau, J. M. (2010). The empirical status of social learning theory: A meta-analysis. *Justice Quarterly*, *27*, 765–802.

Prescott, A. T., Sargent, J. D., & Hull, J. G. (2018). Metaanalysis of the relationship between violent video game play and physical aggression over time. *Proceedings of the National Academy of Sciences, USA*, *115*, 9882–9888. doi:10.1073/pnas.1611617114

Przybylski, A. K., Deci, E. L., Rigby, C. S., & Ryan, R. M. (2014). Competence-impeding electronic games and players' aggressive feelings, thoughts, and behaviors. *Journal of Personality and Social Psychology*, *106*, 441–457. doi:10.1037/a0034820

Przybylski, A. K., & Weinstein, N. (2019a). Digital screen time limits and young children's psychological well-being: Evidence from a population-based study. *Child Development*, *90*, e56–e65. doi:10.1111/cdev.13007

Przybylski, A. K., & Weinstein, N. (2019b). Violent video game engagement is not associated with adolescents' aggressive behaviour: Evidence from a registered report. *Royal Society for Open Science*. doi:10.1098/rsos.171474

Pustejovsky, J. (2019). Sometimes, aggregating effect sizes is fine. Retrieved from https://www.jepusto.com/sometimes-aggregating-effect-sizes-is-fine/#ref-VandenNoortgate2015metaanalysis

Quandt, T., Van Looy, J., Vogelgesang, J., Elson, M., Ivory, J. D., Consalvo, M., & Mäyrä, F. (2015). Digital games research: A survey study on an emerging field and its prevalent debates. *Journal of Communication*, *65*, 975–996.

Quintero-Johnson, J., Banks, J., Bowman, N., Carveth, R., & Lachlan, K. (2014, April). Four respond to 228, responding to the APA: Dissent within the academy regarding media violence. Symposium presented at the Eastern Communication Association, Providence, RI.

Regenbogen, C., Herrmann, M., & Fehr, T. (2010). The neural processing of voluntary completed, real and virtual violent and nonviolent computer game scenarios displaying predefined actions in gamers and nongamers. *Social Neuroscience*, *5*, 221–240.

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

Roth, P. L., Le, H., Oh, I.-S., Van Iddekinge, C. H., & Bobko, P. (2018). Using beta coefficients to impute missing correlations in meta-analysis research: Reasons for caution. *Journal of Applied Psychology*, *103*, 644–658. doi:10.1037/apl0000293.supp

Rothstein, H. R., & Bushman, B. J. (2015). Methodological and reporting errors in meta-analytic reviews make other meta-analysts angry: A commentary on Ferguson (2015). *Perspectives on Psychological Science*, *10*, 677–679. doi:10.1177/1745691615592235

Sakaluk, J. K., Williams, A. J., Kilshaw, R. E., & Rhyner, K. T. (2019). Evaluating the evidential value of empirically supported psychological treatments (ESTs): A meta-scientific review. *Journal of Abnormal Psychology*, *128*, 500–509. doi:10.1037/abn0000421

Sauer, J. D., Drummond, A., & Nova, N. (2015). Violent video games: The effects of narrative context and reward structure on in-game and postgame aggression. *Journal of Experimental Psychology: Applied*, *21*, 205–214. doi:10.1037/xap0000050.supp

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

Savage, J., & Yancey, C. (2008). The effects of media violence exposure on criminal aggression: A meta-analysis. *Criminal Justice and Behavior*, *35*, 1123–1136.

Scharrer, E., Kamau, G., Warren, S., & Zhang, C. (2018). Violent video games do contribute to aggression. In C. J. Ferguson (Ed.), *Video game influences on aggression, cognition, and attention* (pp. 5–21). Cham, Switzerland: Springer.

Schimmack, U. (2014). *Quantifying replicability: The R-index*. Retrieved from https://replicationindex.com/about/

Segev, A., Rovner, M., Appel, D. I., Abrams, A. W., Rotem, M., & Bloch, Y. (2016). Possible biases of researchers' attitudes toward video games: Publication trends analysis of the medical literature (1980–2013). *Journal of Medical Internet Research*, *18*(7), Article e196. doi: 10.2196/jmir.5935

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*, 666–681. doi:10.1177/1745691614553988

Society for Media Psychology and Technology. (2020). [Letter to the APA council of representatives on the resolution on violent video games.] Retrieved from https://www.scribd.com/document/448927394/Division-46-Letter-to-the-APA-criticizing-it-s-recent-review-of-video-game-violence-literature

Sternheimer, K. (2007). Do video games kill? *Contexts*, *6*, 13–17.

Tear, M. J., & Nielsen, M. (2013). Failure to demonstrate that playing violent video games diminishes prosocial behavior. *PLOS ONE*, *8*(7), Article e68382. doi:10.1371/journal.pone.0068382

Tedeschi, J., & Quigley, B. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression & Violent Behavior*, *2*, 163–177.

Tsuji, S., Christia, A., Frank, M., & Bergmann, C. (2019). Addressing publication bias in meta-analysis: Empirical findings from community-augmented meta-analyses of infant language development. *MetaArxiv*. Retrieved from https://osf.io/preprints/metaarxiv/q5axy/

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

Wang, Y., Mathews, V. P., Kalnin, A. J., Mosier, K. M., Dunn, D. W., Saykin, A. J., & Kronenberger, W. G. (2009). Short term exposure to a violent video game induces changes in frontolimbic circuitry in adolescents. *Brain Imaging and Behavior*, *3*, 38–50.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychological journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604. doi:10.1037/0003-066X.54.8.594

Wofford, T. (2015, August 20). APA says video games make you violent, but critics cry bias. Newsweek. Retrieved from http://www.newsweek.com/apa-video-games-violence364394

Zgoba, K. M. (2005). Spin doctors and moral crusaders: The moral panic behind child safety legislation. *Criminal Justice Studies*, *17*, 385–404. doi:10.1080/1478601042000314892