

Are Associations Between “Sexist” Video Games and Decreased Empathy Toward Women Robust? A Reanalysis of Gabbiadini et al. 2016

Christopher J. Ferguson¹ · M. Brent Donnellan²

Received: 31 March 2017 / Accepted: 25 May 2017
© Springer Science+Business Media New York 2017

Abstract Gabbiadini, A., Riva, P., Andrighetto, L., Volpato, C., & Bushman, B. (PloS ONE, 2016) provided evidence for a connection between “sexist” video games and decreased empathy toward girls using an experimental paradigm. These claims are based on a moderated mediation model. They reported a three-way interaction between game condition, gender, and avatar identification when predicting masculine ideology in their original study. Masculine ideology was associated, in turn, with decreased empathy. However, there were no main experimental effects for video game condition on empathy. The current analysis considers the strength of the evidence for claims made in the original study on a sample of 153 adolescents ($M_{\text{age}} = 16.812$, $SD = 1.241$; 44.2% male). We confirmed that there was little evidence for an overall effect of game condition on empathy toward girls or women. We tested the robustness of the original reported moderated mediation models against other, theoretically derived alternatives, and found that effects differed based on how variables were measured (using alternatives in their public data file) and the statistical model used. The experimental groups differed significantly and substantially in terms of age suggesting that there might have been issues with the procedures used to randomly assign participants to conditions. These results highlight the need for preregistration of experimental protocols in video game research and raise some concerns about how moderated mediation models are used to support causal

inferences. These results call into question whether use of “sexist” video games is a causal factor in the development of reduced empathy toward girls and women among adolescents.

Keywords Video games · Empathy · Sexism · Violence · Masculinity

Introduction

The impact of video games on psychological processes has been an issue of contentious debate for decades. These debates have covered numerous areas ranging from potential addiction (Kardefelt-Winther 2015) to cognitive performance (Boot et al. 2011) with much of the attention focusing on issues related to violent content (Ferguson 2015a; Furuya-Kanamori and Doi 2016; Przybylski et al. 2014). Despite many years of research, findings related to outcomes such as youth aggression, crime, and violence have remained mixed. Further, there is little evidence that video games have produced noticeable problems in society at large (Cunningham et al. 2016; Furuya-Kanamori and Doi 2016; Markey et al. 2015b). In recent years, some attention has focused on the potential impact of sexist content in video games. As with most other areas of video game effects, studies in this realm have been mixed with some finding evidence for effects on negative outcomes such as benevolent (but not hostile) sexism toward women (Stermer and Burkley 2015) whereas others have not (Breuer et al. 2015). In 2016, a new study was released claiming to provide further evidence for links between sexist games and decreased empathy toward women

✉ Christopher J. Ferguson
CJFerguson1111@aol.com

¹ Department of Psychology, Stetson University, 421 N. Woodland Blvd., DeLand, FL 32729, USA

² Texas A&M University, College Station, TX 77843, USA

(Gabbadini et al. 2016). However, claims made regarding the study quickly proved to be controversial, with some scholars objecting, both on the PloS One comments and in news media coverage, that the data could not support the relatively strong claims found in the press release (e.g., Singal 2016). The current reanalysis of the original dataset further examines whether the data are supportive of strong claims.

A Developmental Perspective on Empathy

Empathy can be defined as the ability to understand and have sympathy for another's perspective, and reduced empathy is often associated with violations of the rights of others (Jolliffe and Farrington 2004). The development of empathy is complex, involving both genetic and some social factors (Espelage et al. 2004). Normal empathy development appears to unfold along a predictable developmental trajectory, with some signs of empathy evident in the earliest development years and increasing through the teen years and then showing a relatively high degree of rank-order stability across development (Knafo et al. 2008). Clear relationships between this developmental pattern and environmental factors (e.g., bullying victimization) have sometimes been difficult to establish, with inconsistent results (Williford et al. 2016). Genetic influences explain the largest proportion of variance in empathy, with shared environmental influences decreasing across age (Knafo et al. 2008). Although it is not unreasonable to hypothesize that exposure to harsh environments might impair empathy development, not all research has supported this claim (e.g., Frodi and Smetana 1984). This is not to rule out the role of parental socialization in empathy development (e.g., Licata et al. 2016), it is merely to note that empathy development is complex and follows a pattern similar to other complex phenotypes (i.e., larger shared environmental influences early in development that seem to fade with increasing age). The largest environmental effects seem to be found for proximal (e.g., family) rather than distal (e.g., media) influences.

With these issues in mind, the argument for video game exposure in adolescent years having a strong impact on the development of empathy might be less compelling. The development of empathy is complex and profound behavioral changes from short-term exposure to video games would appear to be unlikely. It would not be unreasonable to suggest that exposure to video games with objectionable content might prime thoughts of empathy, but prior evidence from studies of violent video games have not been encouraging. Overall effects of violent game play on player prosocial behavior and empathy, at least for youth, appears to be minimal (Ferguson 2015a). Some studies have suggested that violent game play can actually prime moral

reflection (Grizzard et al. 2014) or more positive views of those who are different when played cooperatively (Adachi et al. 2016). But there does not appear to be a consistent research base to indicate that playing more aggressive games translates to reduced empathy which manifests meaningfully in the real world.

Do “Sexist” Games Promote Decrease Empathy Toward Women?

Several factors must be considered to thoroughly evaluate the issue of whether “sexist” games promoted decreased empathy toward women victims of violence in this study. First, researchers need to have an objective understanding of what constitutes a “sexist” game. Second, researchers must understand the theoretical rationale for why sexist games might promote such outcomes. And third, more specifically, researchers and consumers of research need to fully understand the claims made regarding the original Gabbadini et al. (2016) and how these relate to the previous two issues. We will consider each of these in turn.

What is a sexist game?

There is surprisingly little consensus on what defines a “sexist game” in the literature. Sexist content can generally be considered to include content that emphasizes the physical objectification of one gender (typically women), assigning them inferior status to males (Swami et al. 2010). In reviewing the literature, we found that some studies provided examples (ranging from sexualized content through non-sexual “damsel in distress” tropes) but there was no overarching definition. Although we agree that several examples could constitute sexist content, there seemed to be little clarity on how much of such content was required to make a game into a “sexist game”. Would a game featuring only male characters (such as many of the military shooters) be sexist? Or would Donkey Kong be considered sexist given the “damsel in distress” trope used as the central plot line? This issue is compounded given that some studies allowed survey participants themselves to decide what a sexist game was, without providing much guidance (e.g., Stermer and Burkley 2015).

In experimental studies, the operational definition of a “sexist” game could reflect the scholars’ attitudes toward that game. For instance, Gabbadini et al. (2016) used two entries in the *Grand Theft Auto* (GTA) series (*San Andreas* and *Vice City*).¹ The choice of these games may have been related to the general controversy generated by the GTA franchise. However, it is crucial to note that the GTA games

¹ Both of these games were curiously dated at the time of the study, *San Andreas* having been released in 2004, *Vice City* in 2002.

are sandbox games. This means that players have considerable freedom in the game and the more extreme (and headline grabbing) possibilities are not necessarily an integral part of the experience. Thus, although the GTA does, indeed, have the potential to expose gamers to sexist content, the degree of exposure to such content is somewhat dependent on the choices made by specific players. This could be an issue for a dataset such as the current one where it is *assumed* that players accessed sexist content due to controversies over such content in the game, despite that players may actually have or choose to have only limited exposure to such content during a brief playing session. Accordingly, it may be important to document the integrity of the desired exposure within individual game exposure sessions, although such IV integrity checks are fairly uncommon in the field. Thus, some potential exists for studies, particularly those using games with a sandbox format, to conflate player choices with game experiences and content. The authors noted that the scenes they selected to highlight from the GTA games involved exposure to lap dancers and prostitutes. However, it remains unknown to what degree players remained within these scenes or left to pursue other interests within the game. It is not our intent to be overly critical, but rather to suggest that the degree to which players choose to engage with sexist material may be both more crucial and interesting than randomized exposure conditions themselves. This is particularly true for games such as the GTA series, where players can easily exit areas with sexist content while still remaining within the game.

Why would sexist games promote decreased empathy toward women?

A second issue involves understanding the theoretical mechanisms by which sexist video games would influence empathy toward women. Most theories that examine this issue tend to focus on cultivation of beliefs (Breuer et al. 2015), or social cognitive and/or objectification models of media effects (Stermer and Burkley 2015). Most of these theories work along the line of traditional hypodermic needle model approaches to understanding media effects, which tend to align well with moral concerns about media content (Bowman 2016). Such models generally examine the potential main effects due to exposure to media. Such theories often allow for the potential that some individuals may experience more or less influence due to media. However, at least within the realm of video game effects, some scholars have explicitly claimed that no one is immune to the purported influences of video games. Speaking on the topic in 2014, one of the authors of the Gabbiadini et al. article noted “No one is immune to the effects of violent video games, any more than anyone is immune to the effects of smoking cigarettes” (Savacool

2014). Such comments are not particularly unique to a single article or author (see Markey et al. 2015a for discussion) and this level of generalization might not extend to other areas of media effects. Thus, based on such hypodermic needle models of media effects typically used to frame such studies, main effects due to media exposure should be observed.

Design and Results

An important starting point is to describe the basic design and results of the Gabbiadini study Gabbiadini et al. (2016). The authors used a final sample of 154 Italian high school students and assigned them to play one of three types of video games—neutral games (either a pinball game or puzzle game), violent games (one of two first person shooters), and violent+sexist games (one of two games in the *Grand Theft Auto* series). Although the text of the original articles suggested that participants were randomly assigned to condition, it appears that classes rather than individuals were randomly assigned to game conditions. This may not have been problematic except that the result was that younger participants were overrepresented in sexist game conditions, see Table 1.

The cover story was that researchers were testing cognitive abilities to develop a new video game. The students first watched an introductory video about their game and then played for five minutes under supervision so they could learn to play their game. The main phase of the experiment was a 25-min gaming session set on an intermediate level so that the games were not too easy but also not too frustrating. At the conclusion of the gaming session, students completed manipulation checks, measures of avatar identification, masculine beliefs, and empathy. The empathy dependent variable was based on ratings to eight items capturing responses toward a picture of beaten girl. Students were debriefed about 2 weeks later.

The authors found no evidence of main effects for experimental condition on empathy ($p = .31$). There were effects of experimental condition on avatar identification (scores were higher for the violent games compared to the neutral games and violent+sexist games) and masculine beliefs (scores were higher for the violent+sexist games

Table 1 Crosstabs of age by condition

Game condition	Age					
	15	16	17	18	19	20
Neutral (control)	0	15	14	16	6	0
Violent non-sexist	0	12	20	13	8	2
GTA (sexist)	22	22	3	1	0	0

compared to the neutral games and violent games) ($p_s = .003$ and $.038$, respectively). There was evidence for moderated-mediation using a complex analysis. This analysis involved a three-way interaction between avatar identification, gender and game condition on a mediating masculine role norms (MRNI) variable, with MRNI then predicting empathy. The basic finding was that the association between game condition and masculine beliefs was evident for males who scored relatively highly on the identification variable, which served as the first step of the analysis. In the second step, moving from MRNI to empathy, there was an association between masculine beliefs and decreased empathy ($r = -.348$)². The interpretation was that sexist and violent games were associated with differences in masculine beliefs, but only for males who highly identified with their on-screen avatar. These masculine beliefs were then associated with differences in empathic responding to a picture of a harmed girl.

There also was a main effect for game condition on identification with the avatar (an essential part of the mediation/moderation model ultimately tested); however, the pattern of results was not necessarily straightforward. Avatar identification was highest in the violent games condition. The averages were statistically indistinguishable for the neutral games and violent+sexist games. Thus, there is some indication that games themselves influence avatar identification making the underlying causal model somewhat more complex than what was depicted in their Figure 1. This suggests that the games in the violent condition were potentially more immersive than the games in the other two conditions. This is an interesting observation given that, if one were to argue that more immersive experiences ought to lead to greater effects, it might be expected that less empathy may occur for the violent game group than the sexist game group. However, violent games did not seem to have a causal effect on empathy (and if anything, the average empathy score was actually highest in this condition; see their Table 1).

In addition, the masculine identity variable was negatively associated with reduced empathy across all experimental conditions (e.g., $r = -.321$ in the neutral condition, $n = 51$). Thus, it could be the case that the authors inadvertently capitalized on a pre-existing association between individual differences in masculine beliefs and reduced empathy to infer evidence of a causal chain from the experimental manipulation to reduced empathy for highly identified male players.

To test their mediation/moderation model, the authors used the PROCESS package (Hayes 2013), which is an

efficient and easy to use tool for testing moderation/mediation models using commercially available statistical packages like SPSS and SAS. PROCESS is a regression based approach and provides 76 prepackaged model structures that users can use to test a range of path models. The Gabbiadini article presented a mediation/moderation model (Model 11 in PROCESS) by which the association between game condition (coded so that the neutral condition was -1 , violent was 0 , and violent+sexist was 1) and masculine identity was moderated by avatar identification and gender (that is, there is a three-way interaction for condition, gender, and identification predicting masculine beliefs). Masculine beliefs then served as the mediator between game condition and empathy toward women. Put more simply, the authors tested whether an interaction between game condition, gender and avatar identification influenced MRNI and whether MRNI, in turn, influenced empathy scores.

This multi-part mediation/moderation model was fairly complex and was just one of multiple PROCESS models that might have been tested with the same data. It is difficult to know whether this was an explicitly confirmatory study or whether particular analytic choices were dependent on the data (i.e., the garden of forking paths; see Gelman and Loken 2013). To be clear, we are no way accusing the original authors of any wrong-doing with this suggestion, nor is this type of issue with mediation/moderation analyses unique to the Gabbiadini et al. article. Rather, we are simply raising the possibility that alternative analytic approaches may yield different results. Accordingly, our goal is to test alternative models and evaluate the sensitivity of the results when using different analytic approaches and different measures of the core constructs (available in their dataset). This is important given the accumulating evidence of difficulty replicating some research finding in psychology (Nosek et al. 2012).

The Current Study

The current study is a reanalysis of Gabbiadini et al. (2016) using the original dataset. The intent of this reanalysis is to examine the robustness of the claims about media effects made in light of the effects that would be expected by theory and concerns about the robustness of a three-way-interaction mediation/moderation model. In our reanalysis, we examine the original data analysis, but also consider multiple other PROCESS models that would have been theoretically defensible. Our interest is in examining whether the results reported by Gabbiadini et al. may have been an unintentional artifact of a particular, complex analytic model, or whether these results are robust across multiple reasonable analytic models.

² This was consistent across conditions: Neutral games $r = -.321$ ($n = 51$); Violent games $r = -.299$ ($n = 55$); Sexist games $r = -.429$ ($n = 48$).

As detailed in the Method section, we noticed that the Gabbiadini et al. (2016) dataset generated some puzzling results including an apparent failure of the randomization process. An explanation for this was not immediately clear from the original article, but the authors, in a subsequent comment on the PLoS One site, acknowledged that randomization occurred at the level of classes, not individuals. This resulted in age being conflated with experimental group condition. We are grateful to the authors for this clarification. We acknowledge upfront that we are not merely reanalyzing an existing dataset to highlight how different conclusions might be drawn from the same data. We are also attempting to draw attention to concerns with the dataset itself and to highlight the reality that analytic flexibility may increase the chances to Type I errors. This is instructive both with respect to this study but other similar kinds of studies.

Drawing focus to these issues is important given the attention generated by this article. We believe the evidence for the claims advanced in the original article is weaker than it appears. This is important information for researchers interested in youth development who may wish to build on the original study. In addition, we provide some clear recommendations that could have obviated some of the concerns we raise in this report such as pre-registering experiments and data analytic plans. These procedures may help increase confidence in published results. Illustrating these issues in the context of an existing dataset can prove instructive to other researchers and to consumers of research studies such as journalists who should have a stake in critically evaluating studies for widespread consumption.

Method

Data Acquisition

The original Gabbiadini et al. (2016) manuscript was published in the journal *PLoS ONE*, which requires open data sharing. Thus, the original database was downloaded so we could attempt to reproduce the original results. One covariate (frequency of play) was missing from the dataset. This was graciously provided by the lead author of the

original study upon request. Using that dataset we were able to reproduce the main findings (e.g., we could exactly reproduce the correlation matrix reported in their Table 2).

However, during our examination of the dataset, it appeared that several variables were potentially miscalculated or were calculated in ways that were not well explained in the original manuscript. During personal communication with the lead author, rationales for some of these decisions were provided. In some cases we decided to retain the original scales calculated by Gabbiadini et al. (2016) and in others believed it was more valid to recalculate full scale scores without excluding items as Gabbiadini et al. had done. Where our variables differ from the original variables we explain below (also see Table 2). When we calculated variables, the key dependent variables were based on taking the average of all available responses to each scale. Highly similar results are obtained using the sums.

Participants and Procedure

The original report described the participants as 154 Italian high school students (44.2% male, mean age = 16.812, SD = 1.241, 1 missing gender). One participant did not have empathy data and was excluded from our analyses. The original article reports that the sample was 43.4% male. Participants were assigned to play either a sexist game (*GTA: San Andreas*; *GTA Vice City*) a non-sexist violent game (*Half-Life 1*; *Half Life 2*) or a non-violent game (*Dream Pinball 3D*; *Q.U.B.E. 2*). We note that the non-violent games do not appear to be as carefully matched to the other conditions as would have been ideal, which is a common issue in this field (see Adachi and Willoughby 2011.) In an ideal circumstance, games should differ only on the variable of interest (i.e. sexist content) not on other variables (e.g., frustration, competition, presence or absence of human characters). We recognize that full equivalence between game conditions is difficult to achieve but, in this case, the non-violent games in particular appear to be different in content than either violent or sexist games. The order of the subsequent empathy, MRNI and avatar ID measurements was not specified in the original manuscript.

Table 2 Scales differences between gabbiadini et al. and current reanalysis

Scale	Change from Gabbiadini et al.
Empathy	Recalculated including one potentially dropped item
MRNI	Recalculated with full 15 items available rather than 12 items reported in Gabbiadini et al.
Avatar Identification	Retained original “embodied presence” scale used in Gabbiadini et al. but also analyzed results using the “wishful identification” and “character empathy” scales included in the dataset but not reported in Gabbiadini et al.

MRNI masculine role norms inventory

Materials

Empathy

The main outcome variable was a scale score related to how empathic participants felt toward a picture of an adolescent girl who had been attacked by a boy. The scale score consisted of 8-items with a coefficient alpha of .83. In examining this scale score, it appeared that one item (*affettuosita*) had been dropped from the calculation of the empathy variable used in the published report. We recalculated the empathy variable by taking the mean response to all eight items and this variable was strongly correlated with the original scale in the data file (labelled *emp_scal* in original file: $r = .985$).

Masculine beliefs

To consider the possible mediator variable, masculine beliefs, Gabbiadini et al. (2016) used 12 items from the Male Role Norms Inventory (MRNI; Levant et al. 2010). The MRNI has 39 items and there is a short form with 21 items. The MRNI is typically considered to assess 7 correlated beliefs about masculinity such as self-reliance, the important of sex, toughness, and restrictive emotionality. 12 items were selected to operationalize masculine beliefs for the Gabbiadini et al. study. However, the original Gabbiadini et al. (2016) database has 15 items though 3 had been dropped from the final scale score³. We recalculated the MRNI from the full 15 items ($\alpha = .789$) which was strongly correlated with the reported 12-item scale (labelled *mas_beli* in original file: $r = .968$). The three missing items appeared to be those related to mechanical self-reliance (e.g., “A man should know how to repair his car if it should break down”; Levant et al. 2010).

Avatar identification

The authors used a 6-item “embodied presence” scale to measure avatar identification developed from a scale used by Van Looy et al. (2012). However, unreported in the original article, was the fact that this was one of 3 identification subscales collected by the authors (the other two being wishful identification and character empathy, although Van Looy et al. used different terms for these scales, see below). In personal communication, the lead author explained that

³ During personal communication Dr. Gabbiadini explained that many items for the MRNI were not used because they used language better fit for adults than adolescents or contained language the schools considered explicit. Dr. Gabbiadini also suggested that they may have eliminated the other 3 items that were in the dataset for the same reason. We accepted this explanation as reasonable for using a reduced set of items in the original survey. However, we were not convinced that this logic applied to any decision to drop 3-items from the scale after the data had been collected.

the embodied presence variable best exemplified the degree to which someone personally felt they were the character in the game (sample items from Van Looy et al. 2012 include “I feel like I am inside my character when playing” and “In the game, it is as if I act directly through my character”). We accepted this explanation and retained the original avatar identification (embodied presence) variable ($\alpha = .917$; labelled *avatarID* in the original file). The internal consistency estimate for the five-item wishful identification subscale was .672 which makes its exclusion potentially reasonable (items from Van Looy et al. include “If I could become like my character, I would” and “My character is a better me”). The reliability of the four-item character empathy subscale ($\alpha = .801$) appeared to be reasonably high, so this subscale was also calculated. Nonetheless, it is unclear what content is captured by this variable as Van Looy et al. (2012) describe a six-item similarity identification variable (sample items “My character is an extension of myself” and “I identify with my character) rather than a character empathy scale per se. Why two items were discarded from the original Van Looy et al. scale was not detailed in the original article.

Embodied presence was correlated with wishful identification ($r = .473$) and character empathy ($r = .360$). Wishful identification was also correlated with character empathy ($r = .434$). As we found no difficulties with the original calculation of the Avatar identification scale, the original reported in the data set were used in our analyses.

Covariates

In the original study, manipulation check items for violence ratings and how frequently participants had already played the game differed by condition and were used as covariates. These were retained in the current reanalysis.

Age

During our reanalysis we found that the groups unexpectedly different significantly in the age of participants [$F(2, 151) = 51.480, p < .001$]⁴. Participants in the “sexist” game group were significantly younger ($M = 15.646, SD = 0.699, n = 48$) than in either the neutral group ($M = 17.255, SD = 1.017, n = 51, d = -1.844$) or violent game group ($M = 17.418, SD = 1.100, n = 55, d = -1.923$). Age was used as a covariate in several subsequent analyses. Participant age was also used as a covariate in the original article.

⁴ An examination of crosstabs revealed that all 15-year olds ($n = 22$) were in the GTA condition. Few participants from ages 17–20 ($n = 4$) were in the GTA condition. This grouping of younger participants into the GTA condition appears to be difficult to explain through random assignment of individuals to condition. This result is presented in Table 1. It should have also been noted in the original report given the strong negative correlation between condition in age in their Table 2 (i.e., $r = -.521, n = 155$).

Results

Our analytic strategy is that of a robustness analysis. Given current concerns about *researcher degrees of freedom* (Simmons et al. 2011) which can create unintended false positives due to specific analytic choices that can be dependent on the data, it can be useful to conduct robustness analyses for potential statistical artifacts (Barone et al. 1997). Given that experimenters are presented with an array of potential analytic tools, it is possible that certain analyses may produce biased results that are idiosyncratic to a specific operationalization. When results are robust, the same findings should emerge across the majority of different (but nonetheless reasonable) ways of testing an idea. When results are robust across methods, researchers can have increased confidence in results. If a result is specific to a single approach in cases where other methods would have been theoretical defensible, confidence in the outcome may be reduced. With this in mind, we evaluated the robustness of the original results across several different approaches.

Main Analyses

Note that results are reported here for scales constructed via mean scores. From prior theories on video game effects, we should expect to see a mean difference in empathy toward women in the sexist video game condition compared to the other conditions. This was first assessed using a traditional ANOVA and there was no indication of main effect of video game condition on empathy toward women [$F(2, 151) = 0.881, p = .417, \text{partial } \eta^2 = .012$]. More crucially, the mean empathy rating from the sexist game group ($M = 4.943; SD = 1.155, n = 48$) did not visibly differ from the neutral game group ($M = 4.901; SD = 0.966, n = 51$). Empathy toward women was slightly higher ($M = 5.142; SD = 0.890, n = 55$) in the violent game group, although not significantly so. We then conducted an ANCOVA (gender, age, violence ratings, and frequency as covariates). Results indicated no main effect of video game condition on empathy toward women [$F(2, 144) = 1.880, p = .156, \text{partial } \eta^2 = .025$]. These analyses reproduced the findings reported in the article.

The ANOVA for masculine beliefs calculated with all 15 items was statistically significant [$F(2, 151) = 3.088, p = .049, \text{partial } \eta^2 = .039$]⁵. Means for the MRNI were higher in the sexist game condition ($M = 3.323; SD = 0.800,$

$n = 48$) than the violent condition ($M = 3.103; SD = 0.708, n = 55$) or neutral condition ($M = 2.970; SD = 0.630, n = 51$). The ANCOVA controlling for gender, age, violence ratings, and frequency) indicated a significant game condition effect [$F(2, 144) = 3.588, p = .030, \text{partial } \eta^2 = .047$]. Null hypothesis testing results that are barely significant (i.e., when the p -value is close to .05), can be difficult to interpret, potentially resulting in overinterpretation of “statistically significant” results that are, in fact, relatively weak support for the underlying hypotheses. Bayes Factor Analysis can provide insights regarding the strength of the evidence for a given hypothesis compared to an alternative hypothesis (typically the null in many analyses). Bayes Factor analysis of contrasts between the sexist and neutral condition suggested that the difference is indeterminate, neither definitely supporting the null or alternative hypothesis (BF = 1.77, slightly in favor of the alternative hypothesis). Bayes Factor analysis of contrasts between the sexist and violent condition likewise suggested that the difference is indeterminate, neither definitely supporting the null or alternative hypothesis (BF = 1.79, slightly favoring the null hypothesis). One interesting issue is that the frequency of playing particular games was correlated with masculine beliefs ($r = .226, p < .01, n = 154$). This correlation was similar across conditions albeit not always statistically significant (Control: $r = .182, p = .202, n = 51$; Violent: $r = .275, p < .05, n = 55$; Violent+Sexist: $r = .149, p = .312, n = 48$).

The patterns of results with the covariates as opposed to without the covariates led us to dig deeper and re-examine the covariates, eliminating them one by one. One possible concern is that the presence or absence of a particular covariate might produce $p < .05$ results that are not observed across other more parsimonious analyses. In most cases, eliminating the covariate did not change the outcome when considering statistical significance. However, removing the violence ratings provided by participants as a covariate resulted in a non-significant outcome for game condition on masculine beliefs [$F(2, 147) = 1.358, p = .260, \text{partial } \eta^2 = .018$]. We are uncertain why the presence of the violence rating covariate may have created this effect from a theoretical perspective. Moreover, this rating was intended to serve as a manipulation check (see page 5 of the original) so it unclear whether it is appropriate to include as a covariate. It is also the case that violence ratings are negatively correlated with age ($r = -.238, p < .01, n = 152$) in the whole sample given the apparent failure to successfully randomize participants to condition (Control: $r = .107, p = .458, n = 50$; Violent: $r = .227, p = .095, n = 55$; Violent+Sexist: $r = .244, p = .098, n = 47$). Indeed, there was an overall effect of condition on age [$F = 178.373, df = 2, 149, \text{partial } \eta^2 = .705$].

For the avatar identification construct, the embodied presence variable showed a significant game condition effect was found for both the ANOVA [$F(2, 151) = 5.901,$

⁵ We noticed that one participant (ID# 108) was excluded from these analyses, likely because empathy data was missing for this participant. However, as the empathy data were not necessary for this particular analyses, we also conducted an analysis with #108 included. When this participant was included in the analysis the results were not statistically significant [$F(2, 152) = 2.573, p = .080$] Thus, the outcome of this analysis in terms of statistical significance hinges on the presence or absence of a single participant.

$p < .01$, partial $\eta^2 = .072$] and ANCOVA [$F(2, 144) = 6.989$, $p < .01$ partial $\eta^2 = .088$]. Unexpectedly, embodied presence was highest among the violent game condition group ($M = 4.761$; $SD = 1.212$, $n = 55$) compared to either the sexist game group ($M = 4.122$; $SD = 1.190$, $n = 48$) or neutral group ($M = 3.902$; $SD = 1.585$, $n = 51$) when looking at the means from the ANOVA. This association is not readily detectable from Table 2 in the original given how the game video variable was coded (i.e., as a single linear 1, 0, -1 variable for sexist, violent, non-violent in that order). There was no evidence of a significant condition effect considering the other empathy and wishful identification subscales using either ANOVA or ANCOVA (minimum $p = .150$ for ANOVA for wishful identification).

PROCESS Models

In their original analyses Gabbiadini and colleagues reported results from a PROCESS model (specifically Model 11 using the Process model templates: see <http://afhayes.com/public/templates.pdf>) examining a complex three-way interaction between game condition, avatar identification and masculine beliefs on empathy toward women. This model was not preregistered and the theoretical rationale for this model was not well specified (i.e., why this model was used as opposed to numerous potentially reasonable alternatives). Given that PROCESS allows for the testing of multiple alternate models with numerous variations (76 models are possible in total using the existing templates provided with the software) there is the potential for favoring a particular model that provides evidence consistent with an effect. While running the PROCESS models we used the average responses to the scales to account for missing item response data. Results of the PROCESS models did not differ whether scale sums or scale means were used. PROCESS appears to use listwise deletion by using only those cases that have scores on all variables included in the analysis (see, e.g., <http://www.processmacro.org/faq.html>).

The reported findings were replicated when testing the original model (PROCESS Model 11) from the article using the avatar identification variable (the W variable in Process templates) and with the full masculine beliefs (the M or mediator variable) and empathy scales (the Y or outcome variable). In this model, the pathway from game type (the X variable) to masculine beliefs (M) was moderated by gender (Z) and identification (W). Age, frequency and violence ratings were entered as covariates [1st step: $R = .564$, $R^2 = .318$, $p < .001$; 2nd step: $R = .311$, $R^2 = .097$, $p < .011$]. The effect of the three-way game condition \times avatar identification (embodied presence) and gender was below the $p < .05$ threshold for significance ($t = -2.353$,

$p = .020$). The subsequent path to empathy was likewise significant ($t = -3.391$, $p = .001$).

However, given that the original article used only one aspect of avatar identification (embodied presence) available in the dataset, but not the two other aspects also available (wishful identification and character empathy), this model was estimated with each of these variables in place of embodied presence. Reasonable theoretical arguments could be raised for the use of any of the avatar identification variables, particular wishful identification, which appears most similar to how this construct was used in prior research (e.g., Konijn et al. 2007). As noted by Konijn et al., wishful identification with avatars is generally considered a key component to imitation of avatar behavior. In neither the case of character empathy ($t = -1.104$, $p = .272$) nor wishful identification ($t = -1.347$, $p = .180$) was the hypothesized three-way interaction between game condition, gender and avatar identification statistically significant.

It also seemed reasonable to test several alternative models. First, we tested a basic model (Model 1) in which game condition (the X variable in the PROCESS Model) led to reduced empathy (Y) moderated by masculine beliefs (M). This model was examined as it was the most parsimonious moderation model. We were interested to see if the hypothesized pathways were robust in more parsimonious models not requiring a three-way-interaction. In this model, neither game condition ($t = .161$, $p = .873$) nor the interaction between game condition and masculine beliefs ($t = -.586$, $p = .559$) were significant. Similarly, a variation of this model with the embodied presence avatar ID variable as moderator (M) also was non-significant for either condition ($t = 0.575$, $p = .566$) or the interaction between condition and avatarID ($t = -1.790$, $p = .076$). Lastly a model (Model 7) that excluded the gender variable, which would have been theoretically defensible since theories regarding the effects of video games do not typically hypothesize differential effects by gender, resulted in non-significant outcomes for the interaction between avatar identification (embodied presence) and condition on masculine identity ($t = 1.827$, $p = .0698$). This, of course, differs from more traditional social learning models that emphasize gender specificity in modeling. However, this element of most social learning models hasn't been clarified for video game effects, where clear hypotheses regarding gender specificity have been lacking as part of a general issue of lack of developmental focus (Adachi and Willoughby 2013). Age, frequency and violence ratings were continued as covariates in these models. One might reasonably argue that these findings might be expected if girls were less like to model male models, but previous research and theory in video games has been inconsistent on this issue (e.g., Anderson and Murphy 2003).

We also tested a model in which the roles of avatar identification and masculine role norms were reversed, with avatar identification considered the mediator variable between game condition and empathy, with masculine role norms and gender as moderating variables. This model appeared to be more consistent with prior research (e.g., Konijn et al. 2007) than the original model. This model was specified with Model 11 in PROCESS, just as with the original model, only with the *M* and *W* variables reversed. This model seemed theoretically plausible given that some studies have identified avatar identification as a key mediator variable (Konijn et al. 2007), as opposed to a moderator. This model indicated that game condition ($p = .023$) and gender ($p = .031$) were significant predictors of avatar identification. The three-way condition/gender/male role norms interaction was not significant ($p = .090$) although interactions between condition and male roles norms ($p = .041$) and condition and gender ($p = .045$) were significant. Avatar identification did not subsequently predict empathy ($p = .694$), nor did condition directly ($p = .109$).

Similarly, we tested a model which made better conceptual sense to us than the introduction of a three-way interaction involving avatar identification, considering the relatively low level of avatar identification in sexist games compared to violent games. This model (Model 74) specified the game condition (*X*) predicted the mediator (*M* = masculine beliefs) and moderated the pathway from masculine beliefs to empathy (*Y*). This model, in many respects, appeared to be more in lines with the main study hypotheses than the model tested in the original article. Gender was included as an additional covariate in this model [1st step: $R = .506$, $\Delta R^2 = .256$, $p < .001$; 2nd step: $R = .316$, $\Delta R^2 = .100$, $p = .032$]. In this case, video game condition did not moderate the relationship between masculine beliefs and empathy toward women ($t = -.693$, $p = .490$).

Sexist Content Regression

From our working with the PROCESS models, we became concerned that various outcomes may be contingent on the specific PROCESS models chosen for analysis. Accordingly, we reanalyzed the results evaluating impact of sexist games (dummy coded for GTA vs. the other, non-sexist games) on the empathy outcome. We included both three-way and two-way interaction terms for sexist content with the masculine beliefs and avatar identification embodied presence variables, as well as gender, age, violence ratings, game violence content (dummy coded for the non-violent games against the violent and violent/sexist games) and frequency of play. Masculine beliefs and identification variables were centered prior to creation of interaction terms. Collinearity diagnostics revealed slight collinearity issues (highest VIF was 3.392; attempting interactions

between sexist content and gender created too much multicollinearity) although this was within generally tolerable ranges. Listwise deletion was used for missing data.

This regression model was not statistically significant [$R = .343$, $R^2 = .117$, $F(11, 138) = 1.670$, $p = .086$], likely due to reduced power. Of the included variables only masculine beliefs was significantly associated with empathy toward women ($\beta = -.253$, $p = .027$). All other *p*-values were greater than .10. Likewise, neither the game type by masculine beliefs interaction term nor the game type by avatar identification interaction term was statistically significant. This raises further doubt regarding the degree to which sexist content in games predicts empathy toward women. These results are presented in Table 3.

Note that all of our syntax and outcome files are available at: <http://www.christopherjferguson.com/SGReanalysis.html>.

Discussion

The release of the original Gabbiadini et al. (2016) article generated discussion among scholars regarding the potential role of “sexist” video games in decreasing empathy toward women. With this in mind, the current reanalysis sought to examine that data set to evaluate the robustness of the initial claims.

Our first observation was that randomized game content had no direct impact on empathy toward girls. This is consistent with the reported results in the original article. In contrast to the original authors, however, we regard this as problematic for a causal account that posits that playing sexist video games somehow caused diminished empathic responding in the published experiment. From previous theorizing on video game influences (e.g., Anderson and Murphy 2003), it would be expected that media content should have an overall effect and not just an indirect effect through masculine beliefs (for highly identified male players). This is not to say that some moderation effects are outside the scope of the theory, only that such pathways would typically viewed as enhancing main effects. Accordingly, the absence of main effects is intuitively problematic, particularly for a field that often focuses such unqualified main effects.

We also found that the effect of game content on masculine beliefs was contingent on issues of measurement and model specification. Although we could replicate the original finding, the effect was small and the *p*-value was quite close to .05 ($p = .049$ in the ANOVA we conducted).

Further, in considering the interpretation of the moderated-mediation claim in the original article, it is worthwhile to consider the possibility that this chain capitalized on a pre-existing association between the masculine

Table 3 Regression model for sexist video game content on empathy toward females

Predictor variable	Standardized coefficient	<i>t</i> -value	<i>p</i> -value
Gender	-.014	-0.150	.881
Age	-.137	-1.258	.210
Violence rating	.113	0.730	.467
Frequency of play	.075	0.845	.399
Game violence content	.030	0.204	.839
Game sexist content	-.155	-1.137	.257
MRNI	-.253	-2.232	.027
Avatar ID (embodied presence)	.131	1.331	.185
Sexist × avatar interaction	-.153	-1.526	.129
Sexist × MRNI interaction	.038	0.318	.751
Three-way interaction	-.082	-0.850	.397

Slight collinearity was observed between the violent content dummy code (VIF = 3.392), and sexist content dummy code (VIF = 2.865). Although this did not influence the significance of outcomes, it is possible this may result in slightly inflated standardized regression coefficients. For purposes of meta-analysis, calculations from the main ANCOVA analyses are preferable

MRNI male role norms inventory, Avatar ID avatar identification.

beliefs and empathic responding to pictures of violence against women. This inference is supported by the correlation between these two variables found in the subsample of those high school students who played pinball or a puzzle game (i.e., the neutral game condition which should not have a causal effect on either masculine beliefs or empathy). Statistical analyses involving indirect paths can sometimes capitalize on a “naturally” occurring pathway to produce what seems to be evidence for a causal process. By testing a number of direct and moderated models including subgroup analyses, one might establish a connection between the experimental condition and the putative mediator to “initiate” an indirect effect. Nonetheless, a substantial component of the indirect effect between condition and the outcome variable might simply be a naturally occurring association between the putative mediator and the outcome variable. In this sense, the results from a complex interaction can be indicative of simply a correlation rather than a “real” causal effect due to the experimental IV.

As for the PROCESS mediation/moderation models, although we were able to replicate the original results, we found it to be sensitive to changes in the model. Further the original model tested only one aspect of avatar identification (embodied presence). The authors collected other measures of identification with the avatar but only presented one set of findings. When the model was retested with the character empathy and wishful identification avatar identification variables in place of embodied presence, the key three-way-interaction was no longer significant. As indicated, the use

of these other variables, particularly wishful identification would have been theoretically defensible based on prior research (e.g., Konijn et al. 2007) perhaps more so than embodied presence. Accordingly, the interpretation of the identification idea depends on how variables are operationally defined.

We also have one reservation about the use of sandbox games such as GTA to expose participants to “sexist” content. Although we agree with Gabbiadini et al. that the GTA series has sexist content, given that players have considerable freedom to shape their own experiences, it is harder to determine the strength of the manipulation on each person and the general flow of the causal arrow. Determining that exposure to sexist content is relatively constant in sexist game conditions may be one challenge for this research field in general.

Had the analytic approach and measurement strategy been pre-registered prior to data collection and analysis, the issues we identified in our re-analyses would be far less concerning. In the absence of preregistration, however, it is impossible to know whether analytic decisions were contingent on the results given issues of researcher degrees of freedom (see e.g., Sijtsma 2016; Simmons et al. 2011). To be clear, we are not suggesting any intentional wrongdoing by Gabbiadini et al. (2016) or that our analytic choices are inherently superior to the models published in the article. Our concerns are that researcher expectancy effects may have unconsciously and, in good faith, influenced analysis choices (see also Gelman and Loken 2013). Preregistration helps to alleviate these concerns. Thus, we now turn to a broader discussion of this issue.

The Value of Preregistration

The current case highlights the benefit of preregistered studies (see also Wagenmakers et al. 2012). We note that, in fairness to Gabbiadini et al., preregistration of social science studies remains uncommon. At a most basic level, preregistration means that statistical tests using *p*-values are interpretable. A preregistration plan clearly specifies those analyses that are explicitly confirmatory in nature. Given that expected main effects of game type on empathy were not found in the current data and the claims for “harm” rest on ambiguous indirect effects, we do not think the results provide especially strong evidence for the claims in the article. This is not to say that there may not be reasons to expect interaction effects in some studies and, so long as these are preregistered and based on sound theoretical rationale, these are not necessarily problematic. Nonetheless, preregistration would have reduced concerns about the number of alternative approaches that could be used to analyze these data. Likewise, preregistration would have eliminated concerns that scale construction decisions are

somewhat contingent upon the results as might be the case for masculine beliefs. Researchers can specify how all of the major study variables will be constructed or otherwise specify a decision tree about how measurement decisions will be handled in the preregistration. This step is taken before data are collected or analyzed so none of the decisions are contingent upon the results unless such contingencies are spelled out in advance. Indeed, there are growing concerns about the impact of analytic flexibility on the robustness of research results, especially as they relate to contentious issues such as video games and violence (e.g., Elson et al. 2014). Efforts to spearhead preregistration in media psychology have been promoted particularly by scholars such as Malte Elson, Andrew Przybylski, and James Ivory at various outlets such as a recently released special issue of the *Journal of Media Psychology* on preregistered designs (2016). Fortunately there are now easily accessible methods for preregistration such as AsPredicted.com (<https://aspredicted.org/>) or the Open Science Framework (<https://osf.io/k5wns/>).

We also think it is useful to consider how PROCESS-based studies are evaluated in light of our perspectives on value of preregistration in confirmatory research aimed to inform the public about potential harm related to video games. To be clear, procedures such as PROCESS can be an effective and efficient tool for evaluating pathways of association between variables (i.e., mediation), testing for factors that alter the strength of association between two variables (i.e., moderation), as well as testing factors that moderate pathways (i.e., moderated mediation). PROCESS allows researchers to easily test a wide range of statistical models. The flipside is that this useful tool also affords the opportunity to conduct many analyses in an efficient manner. Thus, we think preregistration is particularly valuable when using PROCESS given the wide range of models that can be tested. Only certain models are theoretically defensible and p -values can be difficult if not impossible to interpret without insight into how many analyses were conducted. Again, we do not believe these problems to be unique to Gabbiadini et al. but rather highlight the need for more rigorous methods across social science.

Accordingly, it would be fruitful if researchers specified their a priori causal model in advance to help constrain the possible PROCESS models that could be evaluated in a confirmatory fashion with a given dataset. For example, the current dataset could be used to test a model whereby GTA games “cause” reduced empathy because they promote increased avatar identification given the freedom of a sandbox game. Alternatively, it might be the case that the causal impact of games with sexist content on empathy is strengthened for those individuals that happen to strongly identify with their avatar (e.g., identification moderates the relation between condition and empathy). These and many

other possibilities seem plausible to us and therefore it would be immensely helpful for authors to pre-register their analyses (and measurement strategy) so a clear distinction can be drawn between confirmatory and exploratory research. We are concerned that PROCESS (like any statistical tool) can be used in an exploratory fashion that ultimately capitalizes on chance to produce findings that do not replicate in confirmatory studies.

Indeed, preregistration is potentially valuable in all fields. Preregistration helps limit (though it probably does not eliminate) situations where researchers consciously or unconsciously make data-contingent decisions thereby increasing the potential that results are specific to only a narrow set of potential analyses that could have been conducted with a given dataset. Some cultural and some practical barriers (such as concerns that publically registered studies may be “stolen”) may limited the use of preregistration in some research areas. However, we note that researchers can keep preregistration plans private for a number of years to avoid being scooped and that many of the objections to preregistration can be addressed. Thus, we believe that there is great value in preregistration.

Given the value of preregistration for helping to constrain analytic flexibility, we think consumers of research should also be altered to the importance of this tool. For example, journalists and other researchers might wish to place more weight in results obtained from pre-registered studies as compared to studies that seem to have a high degree of analytic flexibility. University press offices may also wish to highlight pre-registered studies and place a link to the preregistration in press releases. Journalists can then verify the preregistration and let readers know about this document as well. This will help educate the public about more rigorous scientific methods. Again, preregistration helps to clarify distinctions between confirmatory and exploratory research.

A Developmental Perspective: Why Don't Video Games Influence Empathy?

In one recent meta-analytic review of video game studies conducted with child and adolescent samples, violent video games had little impact on the development of prosocial behavior or empathy (Ferguson 2015a⁶). This meta-analysis examined 101 studies in total, including 18 that examined the impact of violent games on prosocial behavior and empathy. Results indicated that, with other factors controlled, violent video game use had minimal impact on

⁶ The meta-analysis became the focus of lively debate between scholars who supported (Markey 2015) and were critical (e.g., Rothstein and Bushman 2015; Valkenburg 2015) of its conclusions. However, the meta-analysis was independently replicated by Furuya-Kanamori and Doi (2016) who were, in turn, critical of the critiques.

prosocial behavior and empathy. This is consistent with observations that empathy has a fairly typical normative developmental progression that involves a complex interplay between maturational and environmental influences. Moreover, shared environmental influences more pronounced earlier in development and typically reserved for more pervasive and proximal influences such as parenting (Knafo et al. 2008). This early influence for shared environment typically occurs prior to exposure to more action oriented video game play (that is to say, typically within the pre-school years). Accordingly, it would be unexpected for video games to have a narrow influence in a particular facet of empathy development, namely empathy toward women victims of violence, when such games have little influence elsewhere.

The current re-analyses of the Gabbiadini et al. (2016) is consistent with the Catalyst Model of media effects (Ferguson et al. 2008; Surette 2013), which relegates media to relatively minor roles in the developmental process. Outcomes such as the development of antisocial traits (which involves reduced empathy) are viewed as the product of genetic predispositions and parenting, with negative inclinations exacerbated under times of stress. Distal features such as media play limited roles in this developmental process, although they may sometimes provide “stylistic catalysts” that alter the manner in which an antisocial behavior is executed, but not the motivation to commit the behavior in the first place. This approach appears to be consistent with the literature on empathy development (Knafo et al. 2008) in which the development of trait empathy has a large genetic component with shared environmental influences most pronounced during the earliest years of development when exposure to highly violent video games is unlikely.

We note that our reanalysis is concerned only with the issue of sexism in games. Our observations do not necessarily preclude potential problem areas in other areas such as video game addiction (e.g., Wittek et al. 2016). Thus, we hope readers do not misinterpret our conclusions or otherwise over-generalize from the results.

It is worth noting that our reanalysis does not consider other possible confounds that may have influenced study results to create a false positive. For instance, the issue of poor contrasts between game conditions has gotten increasing awareness in recent years and it remains possible that differences between game conditions other than sexist content may have created false positive results (Adachi and Willoughby 2011). It would be ideal to use even more game types in future studies. Furthermore, the study made no mention of a manipulation check for demand characteristics or hypothesis guessing. Given the arguably transparent nature of the study, this may have been another potential confound.

Use of this data set to assert clear and unambiguous causal links between “sexist” games and decreased empathy toward females appears to have been premature. Reanalysis of this data suggest that alternative analytic and measurement approaches can yield more ambiguous evidence than was presented in the article. This is especially problematic in light of the seemingly strong claims made by the researchers in the press release accompanying this article. It is also problematic that there was a systematic association between age and exposure to video games and that the procedures for conducting the study were not described clearly in the article (i.e., participants themselves were not randomly assigned to conditions at the level of individuals).

Conclusion

The issue of whether sexist content in games may influence sexist attitudes and behavior in players has been a topic of recent debate. Prior studies have not conclusively found evidence for a causal relationship between sexist games and sexist players. In the current article, we reanalyzed data from Gabbiadini et al. (2016) to determine how strongly the data support links between sexist games and reduced empathy toward women among adolescents. Our reanalysis raised concerns about the strength of the evidence. Thus, our reanalysis joins an increasing body of literature that suggests there may be little link between sexism in games and sexism in real life. However, this perspective does not mean that moral concerns about sexism in games are unimportant. Our concern is that claims about the power of scientific evidence to support moral agendas may backfire, especially when the evidence is equivocal. Altogether, the current re-analyses highlight the benefits of preregistration, which can offer confidence to research results, particularly those based on fairly complex hypotheses involving moderation and mediation. We hope that this article adds constructively to discussions of video game effects and the broader methodological issues facing all of psychological science.

Funding This reanalysis of a preexisting data set did not have external funding support.

Author Contributions Both authors contributed to the study conceptualization, data analysis and writing of the final draft manuscript. Both authors read and approved the final version of this manuscript.

Compliance with Ethical Standards All procedures described within were developed to comport with APA standards for ethical human participant research.

Conflict of Interest The authors declare that they have no competing interests.

Ethical Approval The original procedures described within received local ethical approval as described in Gabbiadini et al. 2016.

Informed Consent The original procedures described within were conducted with informed consent provided to participants as described in Gabbiadini et al. 2016.

References

- Adachi, P. C., Hodson, G., Willoughby, T., Blank, C., & Ha, A. (2016). From outgroups to allied forces: Effect of intergroup cooperation in violent and nonviolent video games on boosting favorable outgroup attitudes. *Journal Of Experimental Psychology: General*, *145*(3), 259–265. doi:10.1037/xge0000145.
- Adachi, P. C., & Willoughby, T. (2011). The effect of violent video games on aggression: Is it more than just the violence? *Aggression And Violent Behavior*, *16*(1), 55–62. doi:10.1016/j.avb.2010.12.002.
- Adachi, P. C., & Willoughby, T. (2013). Do video games promote positive youth development? *Journal Of Adolescent Research*, *28*(2), 155–165. doi:10.1177/0743558412464522.
- Anderson, C. A., & Murphy, C. R. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior*, *29*(5), 423–429. doi:10.1002/ab.10042.
- Barone, M. J., Shimp, T. A., & Sprott, D. E. (1997). Mere ownership revisited: A robust effect? *Journal Of Consumer Psychology*, *6*(3), 257–284. doi:10.1207/s15327663jcp0603_03.
- Boot, W., Blakely, D., & Simons, D. (2011). Do action video games improve perception and cognition? *Frontiers in Psychology*, *2*, 226 doi:10.3389/fpsyg.2011.00226.
- Bowman, N. D. (2016). The rise (and refinement) of moral panic. In R. Kowert, & T. Quandt (Eds.), *The video game debate: Unraveling the physical, social, and psychological effects of digital games* (pp. 22–38). New York: Routledge.
- Breuer, J., Kowert, R., Festl, R., & Quandt, T. (2015). Sexist games=Sexist gamers? A longitudinal study on the relationship between video game use and sexist attitudes. *Cyberpsychology, Behavior, and Social Networking*, *18*(4), 197–202. doi:10.1089/cyber.2014.0492.
- Cunningham, S., Engelstatter, B., & Ward, M. R. (2016). Violent video games and violent crime. *Southern Economic Journal*, *82*(4), 1247–1265.
- 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*, 419–432.
- Espelage, D. L., Mebane, S. E., & Adams, M. R. (2004). Empathy, caring, and bullying: Toward an understanding of complex associations. In D. L. Espelage, S. M. Swearer (Eds.), *Bullying in American schools: A social-ecological perspective in prevention and intervention* (pp. 37–62). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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). "Everybody knows psychology is not a real science": Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *American Psychologist*, *70*, 527–542.
- Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S., & Smith, S. M. (2008). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation? *Criminal Justice And Behavior*, *35*(3), 311–332. doi:10.1177/0093854807311719.
- Frodi, A. M., & Smetana, J. (1984). Abused, neglected, and non-maltreated preschoolers' ability to discriminate emotions in others: The effects of IQ. *Child Abuse & Neglect*, *8*(4), 459–465. doi:10.1016/0145-2134(84)90027-9.
- Furuya-Kanamori, L., & Doi, S. R. (2016). Angry birds, angry children, and angry meta-analysts: A reanalysis. *Perspectives On Psychological Science*, *11*(3), 408–414. doi:10.1177/17456916166635599.
- Gabbiadini, A., Riva, P., Andrighetto, L., Volpato, C., & Bushman, B. (2016). Acting like a tough guy: Violent-sexist video Games, identification With game characters, masculine beliefs, & empathy for female violence victims. *PloS One*. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152121>.
- Gelman, A., & Loken, E. (2013). The statistical crisis in science. *American Scientist*, *102*, 460–465.
- Grizzard, M., Tamborini, R., Lewis, R. J., Wang, L., & Prabhu, S. (2014). Being bad in a video game can make us morally sensitive. *Cyberpsychology, Behavior, And Social Networking*, *17*(8), 499–504. doi:10.1089/cyber.2013.0658.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta- analysis. *Aggression and Violent Behavior*, *9*(5), 441–476. doi:10.1016/j.avb.2003.03.001.
- Journal of Media Psychology (2016). Technology and human behavior: A preregistered special issue of the *Journal of Media Psychology*. <http://www.icaahq.org/pubs/calls/JournalOfMedia.asp>.
- Furuya-Kanamori, F., & Doi, S. (2016). Angry birds, angry children and angry meta-analysts. *Perspectives on Psychological Science*, *11*(3), 408–414.
- Kardefelt-Winther, D. (2015). A critical account of DSM-5 criteria for internet gaming disorder. *Addiction Research & Theory*, *23*(2), 93–98. doi:10.3109/16066359.2014.935350.
- Knafo, A., Zahn-Waxler, C., Van Hulle, C., Robinson, J. L., & Rhee, S. H. (2008). The developmental origins of a disposition toward empathy: Genetic and environmental contributions. *Emotion*, *8*(6), 737–752. doi:10.1037/a0014179.
- Konijn, E. A., Nije Bijvank, M., & Bushman, B. J. (2007). I wish I were a warrior: The role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Developmental Psychology*, *43*(4), 1038–1044. doi:10.1037/0012-1649.43.4.1038.
- Levant, R. F., Rankin, T. J., Williams, C. M., Hasan, N. T., & Smalley, K. B. (2010). Evaluation of the factor structure and construct validity of scores on the Male role norms inventory—Revised (MRNI-R). *Psychology Of Men & Masculinity*, *11*(1), 25–37. doi:10.1037/a0017637.
- Licata, M., Williams, A., & Paulus, M. (2016). The development of empathy in early childhood. In D. F. Watt, J. Panksepp, D. F. Watt, & J. Panksepp (Eds.), *Psychology and neurobiology of empathy* (pp. 111–148). Hauppauge, NY: Nova Biomedical Books.
- Markey, P. M. (2015). Finding the middle ground in violent video game research: Lessons from Ferguson (2015). *Perspectives On Psychological Science*, *10*(5), 667–670. doi:10.1177/1745691615592236.
- Markey, P. M., Males, M. A., French, J. E., & Markey, C. N. (2015a). Lessons from Markey et al. (2015) and Bushman et al. (2015): Sensationalism and integrity in media research. *Human Communication Research*, *41*(2), 184–203. doi:10.1111/hcre.12057.
- Markey, P. M., Markey, C. N., & French, J. E. (2015b). Violent video games and real-world violence: Rhetoric versus data. *Psychology*

- Of Popular Media Culture*, 4(4), 277–295. doi:10.1037/ppm0000030.
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives On Psychological Science*, 7(6), 615–631. doi:10.1177/1745691612459058.
- 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(3), 441–457. doi:10.1037/a0034820.
- 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(5), 677–679. doi:10.1177/1745691615592235.
- Savacool, J. (2014). Think you're immune to video-game violence? Think again. *USA Today*. <https://www.usatoday.com/story/tech/gaming/2014/03/30/video-games-violence/6437253/>.
- Sijtsma, K. (2016). Playing with data—or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika*, 81(1), 1–15. doi:10.1007/s11336-015-9446-0.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. doi:10.1177/0956797611417632.
- Singal, J. (2016). Here's an overhyped study about video-game violence and misogyny. *New York Magazine*. <http://nymag.com/scienceofus/2016/04/here-s-an-overhyped-study-about-video-game-violence-and-misogyny.html>.
- Stermer, S. P., & Burkley, M. (2015). SeX-Box: Exposure to sexist video games predicts benevolent sexism. *Psychology Of Popular Media Culture*, 4(1), 47–55. doi:10.1037/a0028397.
- Surette, R. (2013). Cause or catalyst: The interaction of real world and media crime models. *American Journal Of Criminal Justice*, 38(3), 392–409. doi:10.1007/s12103-012-9177-z.
- Swami, V., Coles, R., Wilson, E., Salem, N., Wyrozumska, K., & Furnham, A. (2010). Oppressive beliefs at play: Associations among beauty ideals and practices and individual differences in sexism, objectification of others, and media exposure. *Psychology Of Women Quarterly*, 34(3), 365–379. doi:10.1111/j.1471-6402.2010.01582.x.
- Van Looy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012). Player identification in online games: Validation of a scale for measuring identification in MMOGs. *Media Psychology*, 15(2), 197–221. doi:10.1080/15213269.2012.674917.
- Valkenburg, P. M. (2015). The limited informativeness of meta-analyses of media effects. *Perspectives On Psychological Science*, 10(5), 680–682. doi:10.1177/1745691615592237.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 632–638.
- Williford, A., Boulton, A. J., Forrest-Bank, S. S., Bender, K. A., Dieterich, W. A., & Jenson, J. M. (2016). The effect of bullying and victimization on cognitive empathy development during the transition to middle school. *Child & Youth Care Forum*, 45(4), 525–541. doi:10.1007/s10566-015-9343-9.
- Wittek, C. T., Finserås, T. R., Pallesen, S., Mentzoni, R. A., Hanss, D., Griffiths, M. D., & Molde, H. (2016). Prevalence and predictors of video game addiction: A study based on a national representative sample of gamers. *International Journal Of Mental Health And Addiction*, 14(5), 672–686. doi:10.1007/s11469-015-9592-8.

Christopher J. Ferguson is a professor of psychology at Stetson University. His research interests include video game effects and other media effects on behavior, as well as spanking influences and statistical issues in meta-analysis.

M. Brent Donnellan is a professor of psychology at Texas A&M University. His research interests include personality development, self-esteem and narcissism and replication issues in psychology.