I thank the scholars (Boxer et al., 2015, this issue; Gentile, 2015, this issue; Markey, 2015, this issue; Rothstein & Bushman, 2015, this issue; Valkenburg, 2015, this issue) who have taken the time to comment on my “Angry Birds” meta-analysis (Ferguson, 2015, this issue). Debates in this field have historically been unusually acrimonious, owing in part to the extent to which such debates overlap with moral arguments about media content in the general public and policy arenas. Thus it has not often been easy to separate rhetoric from data in this field (Markey, French, & Markey, 2015). The present skew of comments range from the supportive (Markey, 2015) to the self-proclaimed “angry” (Rothstein & Bushman, 2015). What is striking is that, despite much debate, none of the comments fundamentally challenge the conclusion that bivariate relationships between video game use and outcomes are very small and these effects approach zero once controlled multivariate analyses are incorporated. I express the concern at the outset that several of the commenters appear to invite us to ignore these findings, despite having found similar results in several of their own studies, as I will highlight. Although part of this essay will be devoted to addressing some misconceptions and inaccuracies from the comments, I hope also to use this reply to suggest potential ground to move the field forward in a positive, constructive direction.

Selective Reporting/Citation Bias Remains a Serious Issue for the Field

In my original article, I noted that selective reporting/citation bias (literature reviews that fail to cite literature contrary to the authors’ personal views) is a common and serious problem in the field. Further, scholars who engaged in such behavior tended to produce studies with higher...
effect sizes than those who did not. This issue is important to note as the issue seemed to rise up once again in several comments despite my (and others) admonitions of this as problem behavior. For instance Boxer et al. provides numerous “we know” statements about hypothesized video game effects and concludes by saying “We know this from numerous original studies that have controlled for a variety of potential confounds and that still find these effects” (p. 672). Similarly, Gentile states “Several experimental, correlational, and longitudinal studies show that violent video game play predicts these varied types of aggression” (p. 675). Both these statements are examples of selective reportingcitation bias in that they fail to note an increasing number of experimental, longitudinal, and correlational studies that do not find these effects (e.g. Adachi & Willoughby, 2011; Ballard, Visser, & Jocoy, 2012; Ferguson, Garza, Jerabeck, Ramos, & Galindo, 2013; von Salisch, Vogelgesang, Kristen, & Oppl, 2011).

Gentile suggests that, because of limited space, it is not unreasonable for scholars to “tell the story” that best fits their theory. I would counter that scholars are ethically bound to tell the full story, not just the parts that conveniently fit with the narrative they wish to convey. That outside scholars have identified selectively theory-supportive literature reviews as problematic (Babor & McGovern, 2008) bears repeating. I worry also about the degree to which an attitude of “we know” conveys researcher expectancy effects that could influence study results. Boxer et al. fail to convey the inconsistencies and complexities of results regarding not just negative results, but positive ones as well (e.g. van Ravenzwaaij, Boekel, Forstmann, Ratcliff, & Wagenmakers, 2014). Boxer and colleagues also misrepresent my own stance on positive outcomes, referring to an older article and not to more recent articles where I have had difficulty replicating positive effects of video games as well as negative (e.g., Ferguson et al., 2013). Nonetheless, even if video games did have positive influences, positive and negative outcomes must be tested independently.

The commentators also take issue with the observation that applying multivariate controls reduces effect sizes to near zero, but fail to mention several of their own studies that have also confirmed this conclusion. For example, one study coauthored by Valkenburg (Fikkers, Piotrowski, Weeda, Vossen, & Valkenburg, 2013) found that, with proper controls in place, the relationship between violent media including games and aggression in children longitudinally was exactly zero (β = .00). Similarly, a study coauthored by Boxer found an effectively zero effect size (OR = 1.0) for violent video game use on youth violence once other variables were controlled (Ybarra et al., 2008). Thus, in many instances, the authors, in challenging the “Angry Birds” meta-analysis, ignore their own results.

Extreme Claims and Ad Hominem Comments by Scholars

In her comment, Valkenburg states that “What concerns me about the debate between Ferguson and other American media-violence researchers is the tone and the ad hominem arguments that they use” (p. 681). I agree with Valkenburg that scholars must remain cautious about the use of both extreme claims and ad hominem attacks against scholars with whom they disagree (see Ferguson, 2013, for specific cautions against using ad hominem attacks). However, Valkenburg's comment concerns me because it implies that a debate exists between one scholar (me) and other “American” scholars, when, in fact, many scholars from multiple nationalities are involved in both sides of the debate. For example, 238 scholars from multiple nationalities recently wrote to the American Psychological Association asking them to retire their policy statements linking media violence to societal aggression (Consortium of Scholars, 2013).

Debates in the field have certainly been heated. This may, however, be an unavoidable element of paradigm change. Strong criticisms of method, theory, and even public statements are a normal part of the scientific process. Ad hominem attacks have indeed occurred, but, to the best of my knowledge, these have mainly been accusations that skeptics are “industry apologists” (Anderson, 2013) or that they are comparable to Holocaust deniers (Strasburger, Donnerstein, & Bushman, 2014). Even in his comment, Gentile compares scholarly skepticism to “the approach of the cigarette industry” (p. 674).

These statements fit into a larger pattern of extreme statements made by scholars in the field. Some of these statements have involved exaggerations of the magnitude and consistency of results, such as suggesting that media effects are similar in magnitude to smoking and lung cancer (Bushman & Anderson, 2001), claims that the effects of video games are similar to the effects of broken homes (Gentile, as quoted in Almendraia, 2014) or that they are similar in magnitude to those of gang membership and greater than the effects of abusive parenting (Donnerstein, 2014). At other times, causal advocates have claimed a broad consensus among scholars on media effects despite evidence to the contrary (see Fig. 1). Yet more skeptical scholars have cautioned that such statements do more to misinform than inform the public and potentially damage the reputation of our field (e.g., Hall, Day, & Hall, 2011). Table 1 provides some examples of acrimonious statements by media scholars as well as overstatements of the data that appear geared toward shutting down discussion, both in recent and in past years. Markey, Males, French, and Markey (2015) provide a far more extensive list of such statements.
### Table 1. Examples of Acrimonious and Exaggerated/Discussion Chilling Statements by Media Scholars Presently and in Years Past

<table>
<thead>
<tr>
<th>Statements</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad hominem and acrimonious quotations</td>
<td>Strasburger, Donnerstein, and Bushman (2014, p. 572)</td>
</tr>
<tr>
<td>“Despite thousands of research studies on media effects, many people simply refuse to believe them. Some academics may contribute to this because they like to ‘buck the establishment,’ which is an easy way to promote themselves and their research. Of course, many people still believe that President Obama wasn’t born in the United States, President Kennedy wasn’t assassinated, men didn’t walk on the moon, and the Holocaust didn’t occur.”</td>
<td>Boxer (2013)</td>
</tr>
<tr>
<td>“There is no division in the scientific community about this. Despite a vocal minority of scholars, the consensus among scientists, pediatricians and organizations charged with promoting the science of human welfare—is that playing violent video games increases risk for violent behavior.”</td>
<td>Rich (2014, p. 404)</td>
</tr>
<tr>
<td>“Unfortunately, child advocates have sometimes overreached the data, making extreme cautionary recommendations vulnerable to being dismissed as moral panic by industry apologists and academics building careers on contrarianism.”</td>
<td>Huesmann and Eron (2001), responding to a Rolling Stone article by Richard Rhodes critical of media violence research.</td>
</tr>
<tr>
<td>“It is also clear that Mr. Rhodes is not an unbiased observer. He has a conflict of interest when he writes on violence. Not only is his own ego wrapped up in the view that he has been damaged by being abused as a child, but his own financial wellbeing depends on the sales of his book, “Why They Kill” which takes the viewpoint that media violence is unimportant and being violently abused is important.”</td>
<td>Eron and Huesmann (1981, p. 231)</td>
</tr>
<tr>
<td>“Sohn can go back to sleep. His ‘long slumbering doubts about the findings and conclusions of that landmark research…’ need never have been awakened had he taken the time to read more widely in the recent literature…Sohn says he is ‘hard pressed to conceive of a plausible scenario that would explain how the passing of 10 years could result in the growth of the correlation.’ If he would read these reports he would not need to press his intellect so hard.”</td>
<td>Anderson (2013, p. 19)</td>
</tr>
<tr>
<td>“…I hope that the Task Force…successfully distinguish between the true experts and the industry apologists who have garnered a lot of attention with faulty methods and claims.”</td>
<td>Anderson (2013, p. 19)</td>
</tr>
<tr>
<td>Exaggerated magnitude/sensationalistic quotations</td>
<td>Gentile (as quoted in Almendrala, 2014, emphasis added)</td>
</tr>
<tr>
<td>“Violent video games are just one risk factor. They’re not the biggest, and they’re not the smallest. They’re right in the middle, with kind of the same effect size as coming from a broken home.”</td>
<td>Centerwall (1992)</td>
</tr>
<tr>
<td>“Nevertheless, the epidemiologic evidence indicates that if, hypothetically, television technology had never been developed, there would today be 10,000 fewer homicides each year in the United States, 70,000 fewer rapes, and 700,000 fewer injurious assaults.”</td>
<td>Gentile and Anderson (2003, p. 135)</td>
</tr>
<tr>
<td>“If one wanted to learn how to kill someone, one would quickly realize that there are many steps involved. At a minimum, one needs to decide whom to kill, get a weapon, get ammunition, load the weapon, stalk the victim, aim the weapon, and pull the trigger. It is rare for television shows or movies to display all of these steps. Yet, violent video games regularly require players to practice each of these steps repeatedly. This helps teach the necessary steps to commit a successful act of aggression.”</td>
<td>Anderson and Bushman (2001, p. 353)</td>
</tr>
<tr>
<td>“Paducah, Kentucky. Jonesboro, Arkansas. Littleton, Colorado. These three towns recently experienced similar multiple school shootings. The shooters were students who habitually played violent video games. Eric Harris and Dylan Klebold, the Columbine High School students who murdered 13 people and wounded 23 in Littleton, before killing themselves, enjoyed playing the bloody video game Doom. Harris created a customized version of Doom with two shooters, extra weapons, unlimited ammunition, and victims who could not fight back—features that are eerily similar to aspects of the actual shootings.”</td>
<td>Strasburger, Jordan, and Donnerstein (2010, p. 759)</td>
</tr>
<tr>
<td>“The impact of media violence on real-life aggressive behavior is stronger than many commonly accepted public health risks and nearly as strong as the link between smoking and lung cancer.”</td>
<td>Bushman/Ohio State University press release, as cited in Grabmeier (2012)</td>
</tr>
</tbody>
</table>
Why Betas Rule Metas

The commentaries do not focus much on statistical issues, but when they do, the main concern regards the use of partialled effect sizes in meta-analysis. The most strenuous arguments are from Rothstein and Bushman who argue that the Angry Birds meta is “fatally flawed” due to the use of partialled effect sizes (despite my also reporting bivariate effect sizes). Their argument is that the apparent heterogeneous nature of partialled effect sizes makes them unsuitable for meta-analysis and that the results are thus spurious. This argument is problematic both statistically and theoretically.

Statistically, the arguments of Rothstein and Bushman rest upon critical assumptions for which they provide no support, but which probably do reflect some academic folklore surrounding the use of meta-analysis. This folklore specifically involves the notion that bivariate correlations are inherently suitable for meta-analysis, whereas partialled effect sizes are inherently unsuitable. Both of these assumptions can be demonstrated as flawed. Meta-analytic procedures, of course, have no idea whether the effect sizes they are fed are bivariate or partialled. Thus concerns about the use of partialled effect sizes rest on the assumption that partialled effect sizes do not meet with the assumptions for inclusion in meta-analysis, whereas bivariate effect sizes do.

First, it is important to note that meta-analytic procedures for the use of partial effect sizes, including standardized regression coefficients, have been understood for some time (e.g., Paternoster, 1987; Pratt et al., 2010; Savage & Yancey, 2008). Indeed many scholars have argued vehemently that inclusion of partialled effect sizes is superior to bivariate, as I discuss later. However, problems with the assumption that partialled effect sizes are unable to be synthesized can be demonstrated as false as can problems with the assumption that bivariate effect sizes are always suitable for synthesis.

Rothstein and Bushman suggest that the partialling process can result in changes to variance that are unpredictable, potentially producing spurious results. Although they make such claims, they provide no evidence that results from the Angry Bird meta-analysis are spurious. Pratt and colleagues (2010) note that, contrary to Rothstein and Bushman’s suggestions, partialled effects, including standardized regression coefficients, tend to, in practice, demonstrate acceptable elements for incorporation into meta-analysis, including distribution. They also demonstrated that the removal of variance in the partialling process does result in a reduced slope in comparison with bivariate effects but contend this results in more valid estimates than do bivariate correlations. Indeed, larger estimates are not necessarily the most valid estimates. Pratt and colleagues specifically note that the transformation to Fisher’s $Z$ and back to $r$ is appropriate for standardized regression coefficients.

If the variance issues expressed by Rothstein and Bushman were present in the Angry Birds meta-analysis, we might notice this in the distribution of effect sizes. Effect sizes in the Angry Birds meta-analysis were normally distributed, providing initial data against spuriousness. For example, the distribution of aggression/violent

---

**Fig. 1.** Percent agreement among scholars in various fields regarding statements on global warming and media violence effects.
game effect sizes was normally distributed with neither problematic skew (\.63, SE = .30) nor kurtosis (1.98, SE = .58). Although imperfect, the suggestion that the mechanism of converting to Fisher’s Z and then to r may have resulted in spurious effects can be examined by comparing the effect size results to a stripped-down meta-analysis without Fisher’s Z conversion. A stripped down meta-analysis of this nature would actually be considered less accurate, but it would be resistant to the sort of spurious effects that concerned Rothstein and Bushman. Contrary to the concerns of these commentators, a crude averaging of the effect sizes imputed into the Angry Birds meta-analysis for studies of aggression/violent games once again finds an average effect size of r = .07, not substantially different from the Angry Birds results (r = .06). Rothstein and Bushman focus considerably on the heterogeneity statistics, although these are an indication of moderator effects rather than suitability for meta-analysis. Despite their critiques, results for heterogeneity are not much different from the results of their own meta-analysis (Anderson et al., 2010). Rothstein and Bushman critique the Angry Birds meta-analysis for not reporting r or r² despite the fact that they didn’t report these (or F² either) in their own meta-analysis. Had they asked for these statistics, they would have found that these values were low (r = .05 and r² = .003 for aggressive behavior studies), which indicates that although the ratio of heterogeneity due to moderators versus chance indicates the presence of moderators, overall heterogeneity was suitable for meta-analysis. Thus, there is no evidence for Rothstein and Bushman’s claims of spurious effects due to the use of partialled effect sizes.

The assumption that bivariate correlations are better suited to meta-analysis than partialled effect sizes rests on an assumption that bivariate correlations are drawn from a homogeneous pool of effect sizes; using similar methods; testing similar designs; and are free of questionable researcher practices (QRPs), researcher expectancy effects, or publication bias. Pratt et al. (2010) noted that, in most cases, these assumptions are unlikely to be met and that the use of bivariate correlations is often more problematic than better controlled effect sizes. For instance, from a total of 59 separate meta-analyses compiled from 48 papers published in top-tier journals in psychology compiled by Ferguson and Brannick (2012) using either bivariate r or experimental results, 66% exhibited heterogeneity as indicated by tau values above .10 (20% were above .20, with 1 above .30, and 2 above .40). It is not the point that these meta-analyses should not have been conducted, but rather that the issues of heterogeneity and variance are common and may be more problematic rather than less for bivariate correlations.

There are also straightforward theoretical reasons for the preference for controlled rather than bivariate correlations. Most researchers appreciate that it is important to understand what effects remain for a hypothesis once appropriate controls have been applied (e.g., Baumeard, Larzelere, & Cowan, 2002). If meta-analyses remain rooted in bivariate correlations, despite a general understanding that multivariate analyses are superior (Savage & Yancey, 2008), there will be a significant risk of reifying spurious effects (Pratt et al., 2010). Thus, although debates about the usability of partialled effects in meta-analysis is likely to continue for some time (Pratt et al., 2010), I argue that the evidence does not indicate that effect size results are more biased than for bivariate effect sizes and may be, by contrast, more informative.

**Publication Bias**

Currently, publication bias is known to be widespread across psychological science (Kuhberger, Fritz, & Scherdl, 2014). Publication bias was found in the Angry Birds meta-analysis for published studies of aggressive behavior. Nonetheless, the potential biasing effects for publication bias were ignored by most comments or implicitly denied by Rothstein and Bushman. Yet publication bias can be demonstrated not only in the Angry Birds meta-analysis, but in the previous meta-analysis on video games by Rothstein and Bushman (Anderson et al., 2010). For example, examining effect sizes for studies in the best raw analysis of experimental studies of video game violence on aggressive behavior reported effect size correlates with sample size at −.503 (p = .007), a clear indication of publication bias (the negative correlation is −.575 when only published Western samples were considered). Results for the Tandem Procedure also indicate clear publication bias in this group of studies that went unreported in Anderson et al. (2010) with both the rank correlation (τ = .436, z = 3.19, p < .001) and Egger’s regression, t(25) = 4.41, p < .001, tests for publication bias significant, as was the trim and fill (10 studies missing). Thus, results from their own meta-analysis confirm the presence of publication bias. Given the widespread problem of publication bias in psychological science (Kuhberger et al., 2014), it would, indeed, have been remarkable were it not present in video game violence research.

**Continued Methodological Problems in Video Game Research**

I am concerned that several commenters appeared to minimize the extent of methodological problems in video game research. For instance, despite that problems of matching violent and nonviolent games in experiments are now well known (Adachi & Willoughby, 2010), Gentile suggests this has not been a considerable problem, using the work of Craig Anderson as an example. However, a review of Anderson’s work reveals that pretesting was not reliably employed across his coauthored
studies and, when it was, it often found differences between game conditions. Anderson’s work was also among those specifically noted for not matching games on competitiveness, pace of action, and difficulty by Adachi and Willoughby (2010). Thus, failure to find effects for violent content once games are carefully matched (Adachi & Willoughby, 2011; Przybylski, Deci, Rigby, & Ryan, 2014) remains a significant issue.

Gentile also suggests that I made an error in noting the bouncing beta problem in one of his previous papers (Gentile et al., 2009). However, when two variables are highly correlated with each other, yet produce standardized regression coefficients on the same dependent variable in opposing directions when entered together in a regression, this is often symptomatic of multi-collinearity, which can easily occur when the variance inflation factor (VIF) is below 10. Gentile does not tell us the actual VIF value, which might have cleared any misunderstanding. However, the easiest way to understand what may have happened would be for Gentile to make his dataset public and open to scrutiny.

Gentile also questions the need for standardized, clinically validated instruments. Gentile notes outcomes related to fighting, bullying, and so on, but there are many excellent standardized well-validated instruments for exactly these behaviors, such as the Child Behavior Checklist, Olweus Bullying Inventory, and National Youth Survey. There is no reason why scholars need to make ad hoc measures that are unstandardized and potentially open to questionable researcher practices when better instruments already exist. The problem of standardization is also apparent in the measurement of violent video games. For instance, in several of Gentile’s scholarly works using the same dataset, violent game exposure is computed in five different ways:

1. by multiplying self-rated violent content by hours spent playing for three different games and averaging scores (Gentile et al., 2009),
2. by a four-item measure of violence exposure in games with no reliability mentioned (Gentile et al., 2011),
3. by changing the four-item measure to a two-item measure with mean frequency calculated across three games with no involvement of time spent playing (Busching et al., 2013),
4. by a nine-item scale comprised of gaming frequency and three favorite games with violent and prosocial content (Gentile, Li, Khoo, Prot, & Anderson, 2014), and
5. by a six-item scale also comprising gaming frequency, three favorite games, and two-item violent content questions (Prot et al., 2014).

In some studies, the authors do not provide enough information to understand how the video game variables were created and whether violent and prosocial video game questions were treated separately or combined (e.g., Gentile et al., 2014).

A New Road Forward?

Video game research has obviously become very contentious. I would argue this is a natural state of events as an old paradigm (hypodermic needle/social cognitive models implying direct, predictable, universal effects) draws to an end and the need for a new paradigm emerges. At the same time, it is unfortunate that such periods of paradigm shift can be so painful and messy. Thus, rather than end the Angry Birds exchange on a negative note, I would like to offer some positive reflections for the future.

A time for theoretical inclusivity

Much of the controversy over video game research has less to do with data and more to do with the manner in which the hypodermic needle/social cognitive approach was enforced as a kind of absolute truth, as reflected in the comments I selected earlier in this reply (e.g., Table 1). Whether advancing causationist, skeptical, or some self-described “middle ground,” scholars may be quick to assume that only their own views are valid and shut themselves away from differing opinions and data, effectively demonstrating *myside bias* (Stanovich, West, & Toplak, 2013). I consider the selective citation bias issue identified in Angry Birds as one symptom of this, but, to be clear, I do not believe that only scholars on the poles of the debate are prone to this issue; it may be just as problematic among self-described middle-ground views (it is also likely that “middle ground” has little meaning; everyone likely considers themselves middle ground in comparison to someone else).

It may be difficult to separate data from rhetoric in such an environment. So why not call a truce? What would such a truce look like? In essence, we must reach for a culture in which all scholars—causationist, skeptical, and middle ground—open themselves to the possibility of learning from different-minded colleagues. Scholars could reach across ideological divides to find ways to collaborate (e.g., Ferguson & Konijn, in press). In literature reviews, much good will could be generated simply by acknowledging divergent data and opinions and avoiding selective reporting/citation bias in past research. Although it may sound puerile, a kind of Internet “civility pledge” could also be available for those willing to dialogue peacefully, even offering “amnesty” to
those willing to sign on. If we reduce our insistence on rigid adherence to a single theoretical view, one way or another, debates about media effects could be fun rather than angry.

Open science

One way of potentially fixing the methodological debates in the field would be to adopt an open science framework. First, all measurements should be standardized. Second, experimental trials should be preregistered. Third, data could be made public for scrutiny by the scholarly community. At very least, raw data should be submitted to journals during peer review. With greater standardization of methods, transparency, preregistration, and careful scrutiny of data, we may cut down on researcher expectancy effects and questionable researcher practices.

Motivational rather than direct effects

I found much of value in Valkenburg’s comments regarding the notion that a given media might have a very different influence on different people or perhaps no influence on most. So perhaps a violent game might make one person angrier, another calmer, and have no effect on others. At the same time, a given nonviolent game might also make one person angrier, another calmer, and have no effect on others. Put simply, media effects may be hard to quantify in a general direction, but may have much to do with the interaction between the media source, the individual, and what that individual wants to get out of the media. I have previously expressed similar ideas (Ferguson, 2014). If Valkenburg and other scholars are intrigued by the potential for idiosyncratic rather than general media effects, we would have very much to discuss indeed. Unfortunately, research examining this type of model remains scarce (but see Unsworth, Devilly, & Ward, 2007), although this means it may be ripe for further study. Examining media effects from an idiosyncratic approach, particularly one rooted in motivational theory such as self-determination theory (Przybylski et al., 2014), may be instrumental in nudging the field away from the global effects/no effects debate.

Of course some scholars may continue to argue for global effects, and some scholars may continue to be skeptical of most effects—and that should be fine as well. Our field will only truly return to science when it becomes dedicated to open inquiry and dialogue.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Note

1. Rothstein and Bushman refer to this as “idiosyncratic” but do not define what they mean by this. For what it may be worth, both the initial procedure and rebuttals of their criticisms (Ferguson & Heene, 2012) were published in top-tier journals, and the procedure has been used or cited 44 times (as of July 2015) since publication 3 years prior. These observations certainly do not place the Tandem Procedure beyond criticism, only note that it would be helpful if criticisms such as “idiosyncratic” were more specific so that they could be addressed.

References


