
Volume 12, No. 4, July–August 2007
ISSN: 1359-1789
ELSEVIER
AGGRESSION AND VIOLENT BEHAVIOR
A Review Journal
www.elsevier.com/locate/aggviobeh

Available online at www.sciencedirect.com
ScienceDirect
Aggression and Violent Behavior 12 (2007) 470 – 482
AGGRESSION AND VIOLENT BEHAVIOR

Evidence for publication bias in video game violence effects literature: A meta-analytic review

Christopher J. Ferguson *

Department of Behavioral, Applied Sciences and Criminal Justice. Texas A&M International University, Laredo, TX 78045, USA

Received 21 September 2006; received in revised form 25 December 2006; accepted 26 January 2007
Available online 3 February 2007

Abstract

Violence in video games has come under increasing research attention over the past decade. Researchers in this area have suggested that violent video games may cause aggressive behavior among players. However, the state of the extant literature has not yet been examined for publication bias. The current meta-analysis is designed to correct for this oversight. Results indicated that publication bias does exist for experimental studies of aggressive behavior, as well as for non-experimental studies of aggressive behavior and aggressive thoughts. Research in other areas, including prosocial behavior and experimental studies of aggressive thoughts were less susceptible to publication bias. Moderator effects results also suggested that studies employing less standardized and reliable measures of aggression tended to produce larger effect sizes. Suggestions for future violent video game studies are provided.

Keywords: Violence; Aggressive behavior; Computer games

Contents

* Tel.: +1 956 326 2636.
E-mail address: CJFerguson1111@Aol.com.

1359-1789/$ - see front matter
doi:10.1016/j.avb.2007.01.001

The 1999 shooting at Columbine high school during which Eric Harris and Dylan Klebold killed twelve students and a teacher before killing themselves was one of the worst and most infamous school shootings in US history. Exactly what causes a young person (usually male) to engage in a premeditated act of mass-murder against innocent targets is poorly understood. Striking events such as those of Columbine result in a desire among policy makers and the public for ready answers. Interestingly it has become tempting for the public and scientists alike to place responsibility for these crimes on external determinants rather than on the individuals who commit the crimes. Researchers such as Anderson and Dill (2000) readily invoke the memory of Columbine specifically in asserting the relevance of their media violence literature. Although Bushman and Anderson's (2001) claim that media violence is ignored by newspapers and the public as a potential cause of violent crime, Lawrence and Birkland (2004) found media violence (such as the game *Doom*) and other pop culture elements (such as the Gothic subculture) as the 2nd more often discussed cause of the Columbine shooting in articles published in the *New York Times* and *Los Angeles Times*. By contrast less than 1% of news stories focused on the responsibility and moral character of the perpetrators themselves. It thus appears that news outlets may promote media violence in general, and video game violence specifically as a direct cause of violent behavior. Yet it remains unclear whether the scientific evidence provides a compelling case for a direct causal link. Two questions are central to this issue. First, does the size of the effect found in studies of video game violence warrant serious concern, or are they small/trivial effects? And secondly, is there any evidence to suggest publication bias in the research on video game violence effects that may be used to misinform the scientific community and public in regards to the strength of this association? It is to these matters that this paper concerns itself.

1. The research on video game violence effects

Prior to discussing the research on video game effects, it is worth noting that the extant literature often relies on studies which demonstrate "statistical significance." However, statistical significance has been recognized as a poor and biased means of determining the importance of a study's findings (Cohen, 1994; Loftus, 1996). As statistical significance is sensitive to sample sizes, studies with large sample sizes may produce statistically significant effects for data with relatively small or trivial effect sizes. Thus, it is worth considering not only whether a study's results are statistically significant, but whether the effect size is meaningful.

The research on violent video game effects often associates violent video game play with violent criminal behavior. Anderson (2004) begins his meta-analytic review with a discussion of school shootings that have been "linked" to playing video games. In order to support an association between violent video game playing and violent crime, researchers rely on multiple forms of information, hoping to find agreement among them. Like research on violent media effects in general (e.g., Heusmann, 1994), video game violence effects research generally falls into two categories: correlational (or non-experimental) research and experimental, laboratory-based research. If playing violent video games in real life can be correlated with aggressive behavior and similar effects are seen due to random assignment to a violent video game condition in laboratory experiments, it is thought that this provides evidence of a causal link between video game violence and aggressive behavior. The model for this association is explained through the General Aggression Model (Anderson & Dill, 2000), which relies heavily on social modeling theories. Despite the relatively young and sparse nature of the research on violent video game effects, some researchers have claimed that the evidence is conclusive (e.g., Carnagey & Anderson, 2004).

Yet a close read of the literature reveals that many of the studies used to support this link provide only questionable or inconsistent evidence. An example of this is the Anderson and Dill (2000) experimental study which examined the impact of random assignment to a violent or non-violent video game condition on setting a computer noise blast level during a competitive reaction time test (the Taylor Competitive Reaction Time Test; TCRTT). Despite claims by the authors that this measure of aggression has "external validity," no study of the association of this measure with actual criminally violent activity has ever been conducted. The studies used to cite "external validity" (i.e., Anderson & Bushman, 1997; Anderson, Lindsey, & Bushman, 1999; Giancola & Chermack, 1998; etc.) actually provide no evidence that higher use of noise blasts is associated with any external indicator of aggression (such as criminal violence) within a sample of participants. Rather these studies use very indirect methodology to suggest that similarity of results in correlational and experimental studies provides an indication of external validity for the TCRTT. Other studies (e.g. Giancola & Zeichner, 1995) provide evidence of the construct validity of using electric shocks as a measure of aggression, not noise blasts more commonly used in the literature for video game violence effects. For a discussion of the conceptual and methodological concerns with laboratory paradigms of aggression see Tedeschi and Quigley (2000).

Regardless of the TCRTT's validity concerns, the authors (Anderson & Dill, 2000) used four methods of measuring aggression derived from this one task (noise intensity and duration after both win and loss trials) which runs the risk of capitalizing on chance. Only one of the four measures of aggression (noise duration after loss trials) was significantly associated with playing a violent video game. The combined effect size of the four measures was not examined, nor were confidence intervals around the effect size for aggressive behavior. Had this been done, it would have been demonstrated that the effect size for this experimental outcome approaches zero and that the confidence interval for the effect size for negative behavior clearly crosses zero and should not be regarded as positive evidence. Nonetheless, this study is often cited as one of the leading studies indicating a link between video game violence exposure and aggressive behavior in the lab. Examined closely it appears to indicate quite the opposite.

One concern is that the Taylor Competitive Reaction Time test is used to measure aggression differently in various studies of violent video game effects (see Anderson & Murphy, 2003; Carnagey & Anderson, 2005 and Bartholow, Bushman, and Sestir, 2006, for three alternate ways of using the TCRTT). This indicates that there is no standardized way of interpreting the TCRTT which allows haphazard use of the measure and capitalization on chance. Other measures of "aggression" applied in laboratory settings include relatively innocuous physiological measures of arousal such as brainwaves (Bartholow, Bushman, and Sestir, 2006) and heart rate (Fleming & Rickwood, 2001), interpreting the actions of a character in a story (Bushman & Anderson, 2002), sentencing criminals in an analog scenario (Deselms & Altman, 2003) and "aggressive thoughts" (Anderson & Dill, 2000). To date, none of these indices of "aggression" has been linked with actual criminally violent behavior. Further, as with Anderson and Dill (2000) many of these studies employ multiple measures for the same dependent construct and find significance for only some of them. Yet, inconclusive results are often ignored during the subsequent "discussion" suggesting the study as one "confirming" a link between violent video games and aggression rather than acknowledging the inconclusive nature of their findings. Rather than selectively describing "statistically significant" results, it would be more meaningful to collapse effect sizes across related measures (that measure the same dependent variable) and discuss the confidence intervals of the resulting effect sizes.

As to non-experimental research on the link between violent video games and aggressive behavior, the Anderson and Dill (2000) study once again is an often cited manuscript. In addition to the experiment discussed above, this manuscript reports a correlation between exposure to violent video game exposure and self reported aggressive behavior that is actually fairly strong. However, the data from this study fails to take into consideration family background, a variable which may lead both to a preference for violence video games as well as actual aggressive behavior. Further, including a single obvious "face valid" measure of aggressiveness is not adequate to claim that the authors have adequately controlled for personality characteristics (such as psychopathy) that are involved in the commission of violent crime (Hare, 1993).

No other correlational study has found such a strong association between video game playing and actual aggressive behavior as do Anderson and Dill (2000). Many such findings (e.g., Colwell & Kato, 2003; Funk, Buchman, & Germann, 2000; Wiegman and van Schie, 1998) have effect sizes that are close to zero and for whom confidence intervals would likely cross the zero point. Once again, many studies use multiple measures for the same dependent construct with inconsistent results, yet still discuss the overall study as positive rather than inconsistent.

2. Meta-analyses of video game violence and the publication bias issue

Two previous meta-analyses (Anderson & Bushman, 2001; Sherry, 2001) have produced somewhat different results. The Anderson and Bushman (2001) meta-analysis examined studies across multiple outcome variables including laboratory "aggressive" behavior, aggressive cognitions (including affect), prosocial behavior, and physiological measures such as heart rate. Generally, their results support a fairly consistent effect for violent video game exposure with effect size (r) ranging from −.16 for prosocial behavior to .27 for aggressive cognition (aggressive behavior had r=.19). While small these effects represent consistent findings that could be considered important.

However, Anderson and Bushman provide few details on how they addressed several key issues. For example, it is not clear from their report how they handled situations in which multiple independent or dependent measures of the same construct were reported in a study (such as in Anderson & Dill, 2000). Meta-analysis is an imperfect procedure (Bobko & Stone-Romero, 1998) that requires some decisions regarding the manner in which data will be handled. It is helpful to know the decisions made by authors conducting meta-analyses in order to better understand the practical importance of their results. Both meta-analyses which make "liberal" and "conservative" decisions on data handling may be useful, but it is helpful to know which category a particular analysis falls into.

By contrast Sherry's (2001) meta-analysis found somewhat more equivocal results for the relationship between violent-video game playing and aggression. Sherry provides more detail than Anderson and Bushman (2001) about how problems related to multiple measures for the same construct (with differing outcomes) were handled. Sherry found a lower overall effect size ($r = .15$). Sherry noted that these results are considerably weaker than those found in meta-analyses of television violence (e.g. Paik & Comstock, 1994). However, unlike Anderson and Bushman (2001), Sherry appears to have collapsed all categories of findings into one main analysis, thus not allowing for much differentiation of the relative strengths of correlational versus experimental research on varying dependent variables.

3. The publication bias issue

Unfortunately, neither Anderson and Bushman (2001) nor Sherry (2001) provided any analyses of publication bias. As discussed in Rosenthal and Rosnow (1991) publication bias (or the "file drawer effect") occurs when articles with positive (i.e., statistically significant) results are selected for publication to a greater proportion than are articles which report negative results. As a result, the extant literature in peer-reviewed publications may provide a biased sample of all of the studies actually carried out, portraying more positive findings than actually exist. This bias may be passed on to meta-analytic procedures as many of those unpublished "negative" studies would likely be difficult to locate. Arguably this "file drawer effect" may happen at two levels. First, journal editors may prefer "positive" findings for publication as these are more "interesting." Thus manuscripts with "negative" findings are more often rejected for publication than those with "positive" findings. Secondly (and perhaps more cynically) authors themselves may suppress (consciously or unconsciously) research which does not conform to their a priori hypotheses. This may occur either through suppressing an entire study with "negative" results or rerunning statistical analyses using multiple variations until the desired results are produced. This need not imply intentionally unethical behavior as researchers may rationalize the initial results as "mistaken" or due to some initial "error" on their part which, once corrected, produces the expected result.

As Rosenthal and Rosnow (1991) note, there is no perfect means of estimating publication bias. However Rothstein, Sutton, and Borenstein (2005) notes that there are several methods for estimating publication bias. These methods include:

a.) Visual examination of a "funnel plot," in which asymmetrical results are an indication of publication bias.
b.) The Fail-safe N. This technique, as suggested by Rosenthal and Rosnow (1991) involves computing a combined *p*-value for all of the studies included in the meta-analysis, and calculating how many additional studies with a zero effect (average z of zero) would be necessary to create a non-significant p. With robust findings the value of the fail-safe N is usually very high (oftentimes in the thousands).
c.) Orwin's fail-safe N. Orwin (1983) provided an alternate formula for calculating the number of studies necessary to bring the effect size down to trivial levels (e.g., $r \leq .10$). This version of the fail-safe N focuses on "practical significance" rather than "statistical significance." As meta-analyses profit on combining sample sizes to very large numbers, it is quite possible to produce results that are statistically significant, yet nearly meaningless on a practical level.
d.) Begg and Mazumdar (1994) rank correlation test provides a rank-correlation (Kendall's tau) for the relationship between effect size and the standard errors of the effects. Significant results indicate a relationship between effect size and precision. This is a relatively conservative test with low power, thus significant findings are usually a good indication of publication bias (Rothstein, Sutton, & Borenstein, 2005).
e.) Egger's Regression (Egger, Davey-Smith, Schneider, & Minder, 1997). As with Begg and Mazumdar's method, this test attempts to quantify the bias captured in the funnel plot. As this uses actual effect sizes and standard errors, rather than ranking, it is a more powerful test than the rank correlation test.

f.) Duvall and Tweedie's (2000) Trim and Fill. This iterative procedure provides and adjusted estimate of effect size that includes the expected value of missing studies that would create a symmetrical funnel plot. This provides an estimate of what the effect size would be if there was no publication bias in the meta-analysis.

Similar to the way that concordance among multiple "goodness of fit" indices in confirmatory factory analysis can be used to demonstrate overall goodness of fit of a factor structure (see Lance & Vandenberg, 2002, for a discussion), concordance among these various indices of publication bias can provide consistent evidence of this concern in a meta-analysis.

4. The current analysis

The current meta-analytic review is designed to answer questions related to the effect size of non-experimental and experimental studies on the relationship between violent video game exposure and aggressive behavior, aggressive thoughts (cognitions and affect), prosocial behavior, and physiological responses. The current meta-analytic review has three main goals: (a) to provide an estimate of the effect size for the relationship between violent video game exposure and the four dependent constructs mentioned above for research occurring over the previous ten years (1995–2005) that is corrected for publication bias, (b) to examine whether there is any evidence of publication bias in the research reviewed here, and (c) examine any moderator variables that may be associated with differences in effect size noted across studies.

5. Method

This section will describe three aspects of the meta-analytic process: (a) the method for selecting and categorizing studies including in the meta-analysis, (b) the method for calculating effect size estimates from the studies, and (c) the procedure for statistical and publication bias analysis.

5.1. Study selection and categorization

The initial literature search procedure was similar to that employed by Anderson and Bushman (2001). PsycINFO was searched for all articles published between the years of 1995 and 2005 (this criteria discussed below) that included the following search terms: (video* or computer or arcade) and (game*) and (attack* or fight* or aggress* or violen* or hostil* or ang* or arous* or prosocial or help). Note that these search terms are identical to those employed by Anderson and Bushman (2001). The references of primary sources revealed in this search were also examined for studies that were not discovered during this initial search.

Articles were judged relevant if they met the following criteria:

a) Articles had to have been published between the years of 1995–2005. There were two reasons for examining this time-frame. The first was to examine trends in effect size within "recent" research. Secondly, and perhaps more importantly, Carnagey and Anderson (2004) have identified this period (1995–2005) as the "third era" in which video game graphics improved markedly over previous eras, on-line playing has become more common, and first-person shooter type games have increasingly predominated the market. This "third era" is marked by a great increase in the inclusion of violent content in commercial video games. It was felt important that the meta-analysis conducted in this manuscript reflect research on the most current gaming technology, as this "third era" is the period in gaming technology which has caused the most controversy/concern regarding violent effects.
b) Articles had to examine the effect of playing violent video games on some measure of aggressive behavior, aggressive thoughts (cognitions and affect), prosocial behavior or a form of physiological measure such as heart rate. Articles that did not distinguish between violent and non-violent video games were not included in this analysis.
c) As this analysis is concerned with the potential for publication bias in peer-reviewed journals, only manuscripts published in peer-reviewed journals were included in the analysis. Book chapters, dissertation manuscripts, and unpublished manuscripts were not included in the analysis. Although it would be interesting and valuable to consider publication status (published or unpublished) as a moderator variable in the analysis, there was no evident

method for assuring that all relevant unpublished manuscripts could be obtained (including those from unknown authors, or those intentionally or unintentionally suppressed by the authors).

A total of 25 published studies was found that met the above criteria. Of these 14 included some form of experimental manipulation, including a total sample size of 1189. Of these studies 13 included some form of non-experimental information, including a total sample size of 3016.

Articles in the current study were coded into one of four main categories according to their dependent (outcome) variable: (a) aggressive behavior (self-reported or laboratory), (b) aggressive thoughts (broadly defined to include a variety of cognitive and/or affective reactions and self-reports that are not aggressive behaviors, but which are theorized to be associated with aggressive behaviors in some way), (c) prosocial behavior including helping, rewarding or empathic responses, and (d) physiological measures, including brainwaves, heart rate, and blood pressure. Some studies included multiple dependent outcomes. In the case that these dependent outcomes were considered to be indices of unique constructs they were maintained separately for analyses. As each dependent outcome will be analyzed in a separate meta-analysis (similar to Anderson and Bushman's 2001 procedure) there was no risk that this could bias the sample size and inflate the results (studies with multiple measures of the *same* dependent variable are discussed below). Studies were also categorized as either experimental or non-experimental. Once again, some manuscripts (e.g., Anderson and Dill, 2000) included separate non-experimental and experimental studies and these were treated as unique studies. Differing from Anderson and Bushman (2001) and Sherry's (2001) analyses, non-experimental and experimental studies will be analyzed separately.

5.2. Calculating effect size estimates

Pearson's *r*, a flexible and easily interpreted index of effect size, was used as the effect size estimate in this study. Similar to Anderson and Bushman (2001), correlation coefficients were transformed to Fisher's *z*, weighted, averaged, and transformed back to a pooled *r*, denoted r_+. Sherry (2001) calculated a mean effect size across each entire study under the assumption that all analyses were measuring the same construct (aggression). Doing so may fail to note differences in effect for various types of measures (i.e., behavioral, cognitive, etc.). Consistent with Anderson and Bushman (2001), this study calculated mean effect sizes across individual constructs (aggressive behavior, aggressive thoughts, prosocial behavior, physiological measures) because these measures may be conceptually independent. As each of these constructs will be examined in a separate analysis there is no concern of including multiple dependent effects in the same analysis. In the case in which a study reported non-significant results but failed to provide statistical information (e.g., *F*-value) the effect size was calculated using the provided means and standard deviations.

Another issue that arises is that of multiple measures for the same construct occurring within a study (multiple dependent or independent measures). Ideally, if the reliability between the measures were known or reported it would be possible to calculate composite score correlation (Hunter & Schmidt, 2004). However, as this information is often not reported, simple mean correlations were computed. In studies in which both univariate (ex. bivariate correlations) and multivariate (ex. partial correlations or correlations adjusted by beta-weights in a multiple regression) were available, only the latter were included in the meta-analysis, as this provided better indices of the unique shared variance between violent video game exposure and aggression (as opposed to that due to gender, trait aggression, etc.).

Table 1
Meta-analytic results for four dependent constructs in experimental studies

Dependent variable	*k*	N	r_+	95% C.I.	Homogeneity test
Aggressive behavior	5	483	.29 (.15)	(.11, .45)	$\chi^2(4)=15.27$, $p \leq .05$
Aggressive thoughts	12	992	.25	(.11, .37)	$\chi^2(11)=50.23$, $p \leq .05$
Prosocial behavior	3	374	.30	(.06, .52)	$\chi^2(2)=10.76$, $p \leq .05$
Physiological	4	363	.27	(.12, .42)	$\chi^2(3)=0.63$, $p \geq .05$

Note, *k*=number of independent studies; *N*=number of participants; r_+=pooled correlation coefficient (coefficient corrected for publication bias is in parentheses); C.I.=confidence intervals.

Table 2
Publication bias results for four dependent constructs in experimental studies

Dependent variable	FSN	OFSN	RCT	RT	DTTF	95% C.I.	Bias?
Aggressive behavior	36	6	$p \leq .05$	$p \leq .001$	.15	(−.06, .35)	Yes
Aggressive thoughts	167	29	$p \geq .05$	$p \geq .05$	N/A*	N/A*	No
Prosocial behavior	29	9	$p \geq .05$	$p \geq .05$	N/A*	N/A*	No
Physiological	7	8	$p \geq .05$	$p \leq .05$	N/A*	N/A*	No

Note. FSN = Fail-safe N; OFSN = Orwin's Fail-safe N; RCT = significance of Begg & Mazumdar's rank correlation test; RT = significance of Egger's Regression; DTTF = corrected r_+ point value for publication bias from Duval & Tweedie's trim and fill; C.I. = confidence interval for Duval & Tweedie's trim and fill.

5.3. Statistical and publication bias analyses

The Comprehensive Meta-Analysis (CMA) software program was used to fit both random and fixed effects models. Anderson and Bushman (2001) argued that random effects models, while less powerful, allow for generalization to a broader population of studies than do fixed effects models. This argument has merit thus, as with Anderson and Bushman's analyses, only random effects models are presented here. Hunter and Schmidt (2004) also argue that random effects models are appropriate when population parameters may vary across studies, as is likely here. Separate meta-analyses were conducted for experimental and non-experimental designs, and for each of the dependent constructs (aggressive behavior, aggressive thoughts, prosocial behavior, and physiological measures). Thus eight separate meta-analyses were actually conducted. This allowed for an examination of the relative merit of each of these types of studies, as well as for publication bias in each of these.

Publication bias was assessed using the six methods described above. General agreement between the six measures was considered to be evidence for or against publication bias.

6. Results

6.1. Experimental studies

Table 1 presents results for the meta-analyses on experimental studies with aggressive behavior, aggressive thoughts, prosocial behavior, and physiological measures as outcome constructs. Results of this analysis revealed moderate effects for experimental studies on aggressive behaviors and thoughts as well as prosocial behavior and physiological measures in laboratory settings. These effects are actually stronger than those reported by Anderson and Bushman (2001) and are consistent with Sherry's (2001) observation of stronger effects in more recent experiments (those occurring in the last 10 years compared to those prior to 10 years ago). Tests of homogeneity were positive (with the exception of physiological measure outcomes) possibly indicating the presence of moderator variables (in contrast to Anderson and Bushman's study). Hunter and Schmidt (2004) note that this test of homogeneity has low power with a

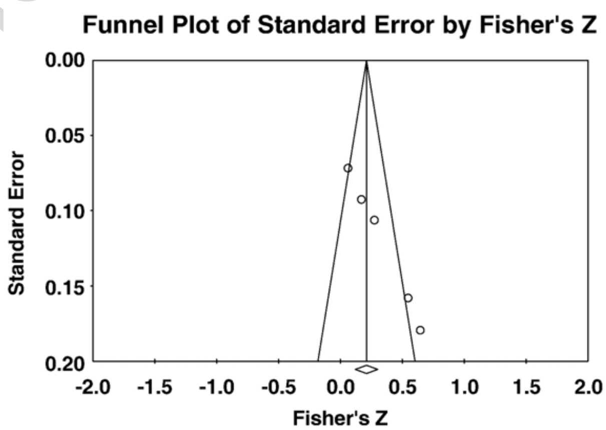

Fig. 1. Funnel plot for experimental studies of aggressive behavior.

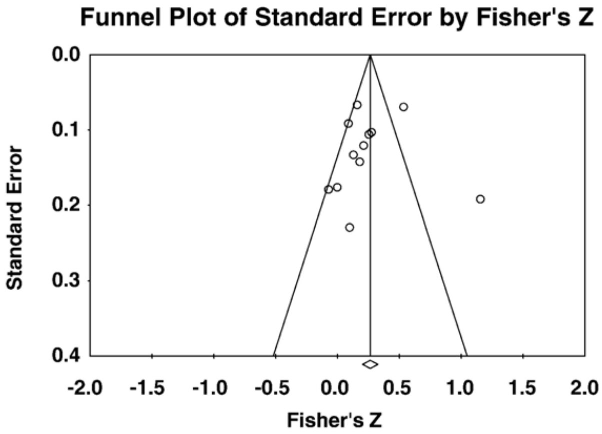

Fig. 2. Funnel plot for experimental studies of aggressive thoughts.

small number of studies, thus a positive finding such as this is a fairly good indicator of moderator variables. Further analyses will look at the qualities (i.e., "best practices") of aggression measure, age of the subject, and year of the study as potential moderator variables.

Of greater concern however, is whether the results provide an indication of publication bias in the literature. Table 2 presents results for the publication bias analyses. Publication bias is indicated by general agreement between the various tests used to measure it. Research on aggressive thoughts and prosocial behavior appears to be generally free of publication bias. However, universal agreement between all measures of publication bias indicates strong publication bias for experimental studies that involve aggressive behavior as the outcome variable. Fig. 1 presents the funnel plot for experimental studies of aggressive behavior, which clearly shows the biased trend (Fig. 2 presents the funnel plot for aggressive thoughts as a non-biased contrast; other non-biased funnel plots were not included). Only 6 additional non-published studies (an almost equivalent number to those published) would render the analysis "trivial." Duval and Tweedie's "trim and fill" adjustment estimates that the "true" effect size is closer to .15, with a confidence interval that includes negative effects. Results for physiological measures were equivocal, with Egger's regression supporting publication bias, but as there was not general agreement between the tests, there can not be said to be clear evidence of publication bias.

Aside from the various indicators of direct publication bias, examination of the fail-safe N and Orwin's fail-safe N indicate the equivocal nature of research in this field. Aside from the experimental association between playing violent video games and aggressive thoughts, which appears to be strong, the experimental association between violent video games and aggressive behavior, prosocial behavior, and physiological measures indicates that only a relatively small number of studies (see Rosenthal & Rosnow, 1991 for a discussion) would render the results as having a trivial effect size.

6.2. Non-experimental studies

Table 3 presents results for the meta-analyses on non-experimental studies with aggressive behavior, aggressive thoughts, and prosocial behavior as outcome constructs (there were no non-experimental studies involving physiological measures). Compared with the experimental effects, correlational effects observed were relatively weak

Table 3
Meta-analytic results for three dependent constructs in non-experimental studies

Dependent variable	*k*	n	r_+	95% C.I.	Homogeneity test
Aggressive behavior	9	2150	.15 (.06)	(.06, .24)	$\chi^2(8)=31.10, p \leq .05$
Aggressive thoughts	5	1067	.13 (.11)	(.06, .20)	$\chi^2(4)=4.50, p \geq .05$
Prosocial behavior	3	771	.13	(.06, .20)	$\chi^2(2)=2.22, p \geq .05$

Note. k = number of independent studies; N = number of participants; r_+ = pooled correlation coefficient (coefficient corrected for publication bias is in parentheses); C.I. = confidence intervals.

Table 4
Publication bias results for four dependent constructs in non-experimental studies

Dependent variable	FSN	OFSN	RCT	RT	DTTF	95% C.I.	Bias?
Aggressive behavior	86	3	$p \le .06$	$p \ge .05$	.06	(−.04, .16)	Yes
Aggressive thoughts	21	2	$p \ge .05$	$p \ge .05$	.11	(.04, .19)	Inconclusive
Prosocial behavior	7	1	$p \ge .05$	$p \ge .05$	N/A*	N/A*	No

Note. FSN = Fail-safe N; OFSN = Orwin's Fail-safe N; RCT = significance of Begg & Mazumdar's rank correlation test; RT = significance of Egger's Regression; DTTF = corrected r_+ point value for publication bias from Duval & Tweedie's trim and fill; C.I. = confidence interval for Duval & Tweedie's trim and fill.

even without correction for publication bias. Only studies involving aggressive behavior as an outcome were significant during the test of homogeneity, possibly indicating moderator effects.

Regarding tests for publication bias, once again studies involving aggressive behavior appear to be subject to publication bias (Table 4). Fig. 3 presents the funnel plot for non-experimental studies with aggressive behavior as an outcome. Egger's regression was not significant, although all other measures indicated a trend toward publication bias. Duval and Tweedie's "trim and fill" indicated that the estimated effect size in the absence of publication bias would be r = .06, with a confidence interval including negative effects. The data on aggressive thoughts was more equivocal (the funnel plot is presented in Fig. 4), demonstrating a mild trend toward publication bias. Egger's regression and the rank correlation test were not significant for this measure, thus there was not a general agreement among the tests, so the results are considered inconclusive.

As with the experimental studies, an examination of the fail-safe N, and Orwin's fail safe N note problems with the strength of the relationship found in these studies. In each case, a relatively small number of unpublished studies would render the results trivial (only 1 study in the case of prosocial behavior).

6.3. Moderator effects

Given that evidence from the tests of homogeneity suggests the presence of moderator variables, several potential moderator variables were tested. Sherry (2001) noted that age of the participant and year of the study acted as moderator variables in their analysis, thus they shall be considered here as well. More critically, the possibility exists that type of instrument used to measure aggression may act as a moderator variable. Specifically it is possible that poorly constructed measures that are used in an unstandardized way may produce greater effect sizes than more standardized ones, given their ability to capitalize on chance or "data fishing." Put more directly, unstandardized (and thus unreliable) measures may be more open to investigator manipulation, wherein the investigator is able to measure aggression in a specific way (out of several possible ways) so as to attain a desired result. In order to address this concern, a "best practices" approach was used to examine for this particular moderator effect. A "best practices" analysis examines whether the use of improper research methodology (such as using unreliable measures) is associated with greater effect sizes. A measure of aggression was considered to be consistent with "best practices" if it reported an

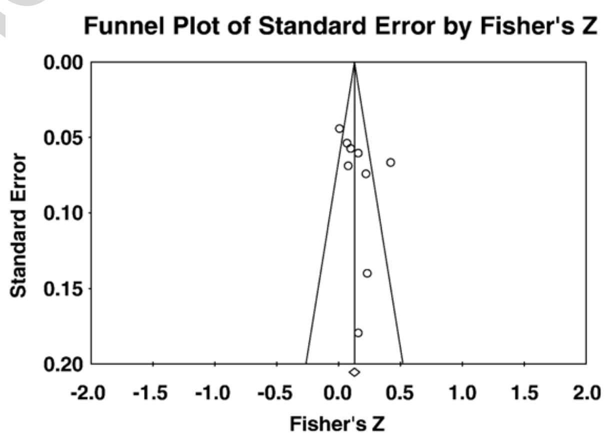

Fig. 3. Funnel plot for non-experimental studies of aggressive behavior.

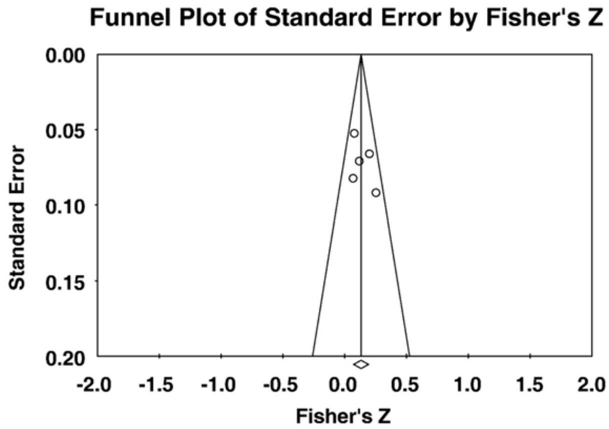

Fig. 4. Funnel plot for non-experimental studies of aggressive thoughts.

adequate (.70 or better) level of reliability and was used in a standardized way consistent with the literature on the development of the instrument. Instruments for which reliability was not reported for the observed sample, or which were used in an unstandardized way across studies (as in the case of the TCRTT) were considered inconsistent with "best practices." Of the 40 individual measurements examined (across studies) only 15 (38%) provided evidence for the reliability of their measures in their study. Best practices were coded dichotomously, with studies reporting adequate reliability and standardized use and those not reporting reliability and standardized use being dummy-coded. If the use of unstandardized, unreliable measures is related to increased effect size, this suggests that the body of literature is being biased by poor methodology. Meta-analytic reviews that do not distinguish between study methodologies may artificially inflate the actual effect size between video game violence and aggressive outcomes.

Consistent with Sherry (2001), age of the participant (r=.29) was found to be a moderator variable, with older participants demonstrating more effects than younger ones. Unlike Sherry (2001) however, there was no evidence that year the study was published (r=−.14) was a moderator variable in this analysis. Although video games are getting more graphically violent (Carnagey & Anderson, 2004) this development in the gaming industry does not seem to be producing greater effects in regards to aggressive behavior. Interestingly, however, year of the study and the use of "best practices" were negatively related (r=−.32), with more recent studies tending to use measures with decreasing standardization. This is an unfortunate trend in the literature.

Related to whether the use of "best practices" instruments used to measure aggression was related to the effect size, results indicated that there was a negative relationship between effect size and "best practices" (r=−.30) measures. Thus unstandardized measures tend to produce greater effect sizes than do standardized and reliable measures. The presence of this moderator variable is of great potential concern for the video game violence literature, and raises the possibility that the actual effect of video game violence on aggressive behavior may have been inflated in the literature.

Because the year of publication and the use of "best practices" instruments were themselves correlated (r=−.32), these moderator variables along with age of the participant were included in a hierarchical multiple regression, with effect size as the dependent variable. The regression equation yielded a multiple correlation coefficient of R=.55 (R^2=.30) which was statistically significant $F(3, 29)$=4.2, $p \leq .01$. In the regression equation, reliability of the aggression measure remained the only predictor of effect size (β=−.44).

7. Discussion

Results of this analysis indicate some concerns for the extant literature on video game violence effects. Publication bias issues emerge for both experimental and non-experimental studies of aggressive behaviors. There also is some indication, although not conclusive, for publication bias in non-experimental studies of aggressive thoughts. Although other areas of research (i.e., experimental studies on aggressive thoughts, prosocial behaviors and physiological measures; non-experimental studies of prosocial behavior) appear to be more sound, an examination of the fail-safe N, and Orwin's fail-safe N reveal that a relatively small number of unpublished or suppressed studies would render the results of this meta-analysis insignificant and/or trivial. As such this meta-analysis indicates that the extant literature on video game violence effects conducted during the “third” era of video games (Carnagey & Anderson, 2004) has not provided compelling support to indicate either a correlational or causal relationship between violent game play and actual aggressive behavior. Moderator analyses also suggested that unstandardized measures of aggression are likely to produce greater effect sizes than are standardized and reliable measures of aggression, a serious concern as the majority (62%) of studies in this analysis fail to conform to guidelines for standardization and reliability for their measures of aggression.

Observing that laboratory results for “aggressive thoughts” are much stronger than for “aggressive behavior” is not unexpected. It comes as no surprise that individuals just exposed to a violent video game in a laboratory should be thinking aggressive thoughts. The important question is whether these “thoughts” then transfer to aggressive behaviors. For example, an individual exposed to a scene depicting suicide is highly likely to be thinking thoughts related to suicide topics, but will this necessarily increase the likelihood of that individual committing suicide, particularly if he or she had not been previously considering it? Even if the existing laboratory measures of aggression were valid (a study as of yet not undertaken) the published results for these studies are a biased sample and thus, the actual result remains unknown. Noting that the effects for correlational studies (even for aggressive thoughts) are uniformly weak/trivial questions the extent to which exposure to video game violence actually is associated with any significant long-term effects.

Part of the problem may be that video game researchers have adopted unreliable methodologies from media violence research in general (Tedeschi & Quigley, 2000). Most of the research (particularly laboratory research) employs unvalidated ad-hoc measures of “aggression.” Perhaps the most often used is the modified Taylor Competitive Reaction Time Test discussed earlier in this paper (which is distinct from its unmodified electroshock predecessor). The TCRTT has never been subjected to validation studies and more pertinently, no reliability data exists for this measure. Most frequently, the participants in these studies are college students or non-pathological samples of students. No experimental studies to date have been conducted on large representative samples of actual violent offenders. A variety of theoretical problems exist with the video game (and media violence) literature, although these problems are beyond the scope of this paper (see Ferguson, 2002 for a brief note).

In order to strengthen the literature on video-game violence effects the following suggestions are offered:

a) Tools used in media violence research need to be properly standardized and empirically validated for reliability and validity. The modified TCRTT is employed differently in almost every study conducted with it, and provides multiple possible ways of measuring “aggression.” No reliability, or validity data exist for this or other laboratory measures of aggressive behavior. The reaction of video-game violence researchers to these concerns (such as those raised by Moeller, 2005) is a surprising “we’ve heard it all before” (see Anderson & Heusmann, 2005), yet the problems remain unaddressed.
b) Research conducted on individuals who actually commit violent crimes would be more productive than that on college students or healthy children. Research conducted on college students and healthy children certainly has merit and can be valuable. However, college students and healthy children represent samples of convenience in which the construct of interest (i.e. violent criminal behavior) is very low. Studies conducted on these populations need to be more direct in addressing the limits to which these studies can be generalized to the issue of violent crime in our society. Although Mook (1983) notes that not all laboratory studies need have “external validity,” they do if the authors intend to generalize their findings “externally” to the “real world” as video-game and other media violence researchers often attempt to do. A few studies do conduct research on children with mental illnesses (e.g., Funk et al., 2002) although such researchers need to be careful not to confound populations of children who are criminally violent with those who have mental illnesses, but are not criminally violent.
c) When multiple measures of aggression (or other variables) are used, interreliabilities between these measures should be reported. In studies such as Anderson and Dill (2000), wherein multiple measures of “aggression” are employed, it would be helpful to know if there is general agreement between those measures (as does not seem to be the case in Anderson and Dill). If there is no general agreement then clearly they are not all measuring the construct of aggression (and without validity studies perhaps none of them are). Even were two measures to measures separate dimensions of aggression, they should, at minimum, be correlated with a sizeable overlap of variance.
d) Researchers need to be more careful in taking account of possible “third” variables that may account (i.e., cause) both violent video game (and other media) consumption and violent behavior. Researchers have generally failed to do this in previous studies. No study to date considers exposure to violence in the family in any regression model (in

many cases, simply adding gender to a regression equation greatly reduces the unique variance attributable to video games). Researchers often claim that they have taken "trait aggression" into account in their studies. However measures of trait aggression such as the Aggression Questionnaire (Buss & Warren, 2000) have very high "face validity" meaning that it is easy for the reader to interpret the intent of the questionnaire. Given that most individuals prone to violent behavior are also prone to lying (Hare, 1993), assertions that personality traits that predispose an individual to violent behavior (such as psychopathy; Hare, 1993) have been controlled seems naïve at best.

The search for video game violence effects is a reasonable one. However researchers must be prepared to test their assumptions and the quality of the data that they are producing. When tragedies such as the Columbine High School shooting occur, it is tempting to look for "scapegoat" answers to a complex problem. It is incumbent on researchers that they not let themselves be side-tracked by a priori hypotheses that may distract the scientific community and the general public from the real biological, social and family influences on violent behavior. Results of the current study raise the concern that researchers in the area of video game studies have become more concerned with "proving" the presence of effects, rather than testing theory in a methodologically precise manner.

References

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

Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology*, *1*, 19−41.

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

Anderson, C., & Dill, K. (2000). Video games and aggressive thoughts, feelings and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, *78*, 772−790.

Anderson, C., & Heusmann, R. (2005). The evidence that media violence stimulates aggression in young viewers remains 'unequivocal'. *APS Observer*, *10*, 7.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, *8*, 3−9.

Anderson, C., & Murphy, C. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior*, *29*, 423−429.

Bartholow, B., Bushman, B., & Sestir, M. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology*, *42*, 532−539.

Begg, C., & Mazumdar, G. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088−1101.

Bobko, P., & Stone-Romero, E. (1998). Meta-analysis may be another useful research tool but it is not a panacea. In G. Ferris (Ed.), *Research in Personnel and Human Resources Management*, *Vol. 16* (pp. 359–397). Greenwich, CT: AI Press.

Bushman, B., & Anderson, C. (2001). Media violence and the American public. *American Psychologist*, *56*, 477−489.

Bushman, B., & Anderson, C. (2002). Violent video games and hostile expectations: A test of the General Aggression Model. *Personality and Social Psychology Bulletin*, *28*, 1679−1686.

Buss, A. H., & Warren, W. L. (2000). *Aggression Questionnaire Manual.* Los Angeles: Western Psychological Services.

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

Carnagey, N., & Anderson, C. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition and behavior. *Psychological Science*, *16*, 882−889.

Cohen, J. (1994). The Earth is Round ($p<.05$). *American Psychologist*, *49*, 997−1003.

Colwell, J., & Kato, M. (2003). Investigation of the relationship between social isolation, self-esteem, aggression and computer game play in Japanese adolescents. *Asian Journal of Social Psychology*, *6*, 149−158.

Deselms, J., & Altman, J. (2003). Immediate and prolonged effects of videogame violence. *Journal of Applied Social Psychology*, *33*, 1553−1563.

Duvall, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89−99.

Egger, M., Davey-Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal*, *315*, 629−634.

Ferguson, C. J. (2002). Media violence: Miscast causality. *American Psychologist*, *57*(6–7), 446−447.

Fleming, M., & Rickwood, D. (2001). Effects of violent versus nonviolent video games on childrens' arousal, aggressive mood and positive mood. *Journal of Applied Social Psychology*, *31*, 2047−2071.

Funk, J., Buchman, D., & Germann, J. (2000). Preference for violent electronic games, self-concept and gender differences in young children. *American Journal of Orthopsychiatry*, *70*, 233−241.

Funk, J., Hagan, J., Schimming, J., Bullock, W., Buchman, D., & Myers, M. (2002). Aggression and psychopathology in adolescents with a preference for violent electronic games. *Aggressive Behavior*, *28*, 134−144.

Giancola, P. R. & Chermack, S. T. (1998). Construct validity of laboratory aggression paradigms: A response to Tedeschi and Quigley (1996). *Aggression and Violent Behavior*, *3*, 237–253.

Giancola, P. R., & Zeichner, A. (1995). Construct validity of a competitive reaction-time aggression paradigm. *Aggressive Behavior*, *21*, 199−204.

Hare, R. (1993). *Without Conscience.* New York, NY: Guilford Press.

Heusmann, L. R. (1994). *Aggressive behavior: Current perspectives.* New York, NY: Plenum Press.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, CA: Sage.

Lance, C., & Vandenberg, R. (2002). Confirmatory factor analysis. In F. Drascow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 221−224). New York: Jossey-Bass/Pfeiffer.

Lawrence, R., & Birkland, T. (2004). Guns, Hollywood and school safety: Defining the school-shooting problem across multiple arenas. *Social Science Quarterly*, *85*, 1193−1207.

Loftus, G. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*, *5*, 161−171.

Moeller, T. (2005). How 'unequivocal' is the evidence regarding television violence and children's aggression? *APS Observer*, *18*, 6.

Mook, D. (1983). In defense of external invalidity. *American Psychologist*, *38*, 379−387.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157−159.

Paik, H., & Comstock, G. (1994). The effects of television violence on anti-social behavior: A meta-analysis. *Communication Research*, *21*, 516−546.

Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research: Methods and data analysis.* New York, NY: McGraw Hill.

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* Hoboken, NJ: Wiley.

Sherry, J. (2001). The effects of violent video games on aggression: A meta-analysis. *Human Communication Research*, *27*, 409−431.

Tedeschi, J., & Quigley, B. (2000). A further comment on the construct validity of laboratory aggression paradigms: A response to Giancola and Chermack. *Aggression and Violent Behavior*, *5*, 127−136.

Wiegman, O., & van Schie, E. G. M. (1998). Video game playing and its relations with aggressive and prosocial behavior. *British Journal of Social Psychology*, *37*, 367−378.