Abstract

Previous research suggests that acute exposure to violent video games are a cause of aggressive behavior. We provide a strong test of this hypothesis by using violent and nonviolent games that are closely matched, collecting a large sample, and using a single preregistered outcome. 275 male undergraduates were randomly assigned to play a first-person shooter game, modified to be either violent or non-violent, difficult or easy. Following the gameplay session, participants were provoked by a confederate and given an opportunity to aggress. Neither game violence nor game difficulty predicted aggressive behavior. Incidentally, we found that 2D:4D ratio, thought to index prenatal testosterone exposure, did not predict aggressive behavior. Results do not support acute violent game exposure and low 2D:4D as causes of aggressive behavior.

Violence is common in video games, and many are concerned about the effects of such games on their players. Psychological research reports that violent games may increase aggressive behavior, and academic societies have made public statements on the harmful effects of violent media (American Psychological Association Task Force on Violent Media, 2005). However, this research has been criticized on the grounds that violent and nonviolent games used in experiments may feature other confounding differences (Adachi & Willoughby, 2011), that outcomes are flexibly analyzed post-hoc to find a significant result (Elson, Mohseni, Breuer, Scharkow, & Quandt, 2014), and that the published literature overrepresents statistically significant results (Hilgard, Engelhardt, & Rouder, 2017).

In this study, we address these challenges. To address potential confounds, we controlled our stimuli by customizing video games. In one version of our game, participants had to kill invading aliens; in the other, participants had to save aliens that happened to be lost. Saving an alien required the participant to transport it back home by aiming a remote controller at it, reproducing first-person-shooter gameplay but without the violent intent. The gameplay in both games was exactly the same; they differed only in graphics, sounds, and cover story. To address the critique that previous studies were underpowered and may have involved post-hoc selection of outcome variables, we collected data from 446 participants, preregistering our sample size, manipulation, hypotheses, and outcome. As a secondary goal, we test whether the ratio of the lengths of the index and ring fingers (2D:4D ratio), believed to measure prenatal testosterone exposure, predicts aggressive behavior as theorized.

**Violent Video Games**

Evidence for causal effects of violent video games comes from laboratory experiments. In such experiments, researchers randomly assign some participants to play a commercially-available violent video game, say *Doom*, and others to play a commercially-available nonviolent video game, say *Myst* (Anderson & Dill, 2000). Following game play, researchers administer some measure of aggressive thoughts, feelings, or behavior. Dozens of such studies have been performed, and meta-analyses of these experiments reveal greater levels of aggression following violent, as compared to nonviolent, video game play (Anderson et al., 2010; Greitemeyer & Mügge, 2014).

This evidence is controversial for two reasons. First, it is often unclear whether observed effects are caused by video games’ violent content in specific. An alternative explanation is that these effects may reflect confounded characteristics of violent video games such as competition or frustration rather than violent content per se (Adachi & Willoughby, 2011; Przybylski, Deci, Rigby, & Ryan, 2014). Second, evidence for violent-video-game effects may be overstated through publication bias (Ferguson & Kilburn, 2010; Hilgard, Engelhardt, & Rouder, 2017).

Violent video games are hypothesized to cause increases in aggression through a number of causal pathways. These include the activation of aggressive thoughts, the operant and observational learning of aggressive scripts, increased processing of ambiguous cues as hostile, desensitization to suffering through repeated exposure to violence, increased arousal, and activation of hostile affect (Bushman & Anderson, 2002). Reported effect sizes are consistent with typical effect sizes in social psychology (*r* = .21, Anderson et al., 2010; *r* = .19, Greitemeyer & Mügge, 2014) and are considered practically meaningful based on their implications for public health.

**Difficult Video Games**

Researchers have attempted to test the specific effects of violent game content without confounding by other game features. Some have suggested that, despite these efforts, differences in violent content between games remain confounded by differences in competitiveness or frustration. One small-sample study suggests that differences in aggression may be attributable to competitive, rather than violent, content (Adachi & Willoughby, 2011; but see Anderson & Carnagey, 2009). Another series of studies reports that frustration with controls, but not game violence, may cause aggressive behavior (Przybylski et al., 2014). These confounds, rather than the violent content, may cause increases in aggression.

**Manipulating Game Content with Better Controls**

Most research manipulates violent content by assigning participants to play a violent or nonviolent game. However, violent and nonviolent games are often very different, usually belonging to very different genres with very different rules of play. For example, violent games are often shooter or fighting games, while nonviolent games are often racing, puzzle, or sports games. Therefore, while tested games do differ in their *violent content,* they are also different in their gameplay, creating a possible confound*.*

Researchers have attempted several ways to account for these potential differences. One approach is to collect a small pilot sample and show that there is no significant confound between games; this approach is flawed in that small pilots do not provide strong evidence of equivalence (Hilgard, Engelhardt, Bartholow, & Rouder, 2017). Another approach is to adjust for potential confounds as covariates. This approach may under-adjust if the confounds are measured with error or overadjust if the “confounds” are themselves consequences of violent game play.

To balance our stimuli, we take a more direct approach by modifying the content of a single video game. Rather than comparing two separate games, game modification allows the researcher to exercise control over the game contents. For example, a game can be modified so that the same level is played either with or without violent content, but all other game parameters are held constant (as suggested by Elson & Quandt, 2016 and demonstrated in Carnagey & Anderson, 2005; Elson et al., 2015; Przybylski et al., 2014). This approach allows manipulation of specific game features in much the same way that a researcher would manipulate features of a laboratory paradigm between conditions, permitting clearer inferences concerning the effects of the manipulated game feature.

**2D:4D Ratio**

While running the experiment, we took the opportunity to test a related hypothesis about hormones and aggression. The male sex hormone testosterone is theorized to be one cause of aggression (see Carré, McCormick, & Hariri, 2011 for a review), and it is hypothesized that development of aggressive tendencies may be caused, in part, by prenatal testosterone exposure (see, e.g., Cohen-Bendahan, Buitelaar, van Goozen, Orlebeke, & Cohen-Kettenis, 2005). One supposed index of this prenatal exposure is the ratio of the lengths of the index and ring fingers (2D:4D ratio); this ratio is thought to be related with both prenatal testosterone exposure and aggressive behavior (see Lutchmaya, Baron-Cohen, Raggatt, Knickmeyer, & Manning, 2004; Manning, Scutt, Wilson, & Lewis-Jones, 1998). There is cause for skepticism, however; meta-analysis indicates that gene polymorphisms that reduce androgen sensitivity do not predict higher 2D:4D as theorized (Voracek, 2014).

Because the overall correlation between 2D:4D ratio and aggression is small (*r* = -.06 among males and no effect among females, Hönekopp & Watson, 2011), proponents of the 2D:4D ratio hypothesis of aggression have suggested the effects of 2D:4D may be moderated by context, demonstrating greater prediction of aggressive behavior in aggressive situations (Millet, 2011). For example, previous experiments report prediction of aggression by 2D:4D following an aggressive (but not a non-aggressive) music video (Millet & Dewitte, 2007) or priming by aggressive (but not neutral) words (Millet & Dewitte, 2009). Our experiment features a provocation, aggressive primes, and an opportunity to aggress, providing an ideal test of the correlation between 2D:4D ratio and aggression.

**Superadditive Causes of Aggressive Behavior**

Several models of aggressive behavior suggest that multiple coincident causes of aggression should produce superadditive interactions (e.g., I3 Theory, Slotter & Finkel, 2011; the General Aggression Model, Anderson & Dill, 2000). For example, one might expect that effects of violent primes are greatest on individuals already temperamentally disposed towards aggression. In this study, we test for interactions between violent content, difficult content, and 2D:4D ratio in predicting aggressive behavior.

**Purpose**

This study examined the effects of game violence, game difficulty, and 2D:4D ratio on aggressive behavior among college-aged males. These effects can be summarized as four hypotheses. H1: Violent video game content will increase aggressive behavior. H2: Video game difficulty will increase aggressive behavior. H3: More masculine 2D:4D ratios will be associated with more aggressive behavior. H4: These effects will yield superadditive interactions.

**Method**

**Participants**

Participants were 446 male undergraduate students at a state university in the American Midwest. Our sample size was planned around a power analysis of the violent-game effect reported in previous meta-analysis, δ = 0.43 (see Anderson et al., 2010). To test this effect with 99% two-tailed power would require 400 subjects. We anticipated a loss of about 50 subjects to failures of methods or deception and set our target size at 450. Power to detect higher-order interactions is harder to estimate, as it is unclear what effect size to expect. Our planned sample size of 400 would have been sufficient to detect effects as small as |ρ| = .12, two-tailed, with 80% power. Previous studies of 2D:4D and aggression have reported simple slopes of *r* = .2-.5 under aggression-promoting circumstances (e.g., Millet & Dewitte, 2007, 2009).

Participation was restricted to males because 2D:4D effects are thought to apply only to males (McIntyre et al., 2007, but see Millet & Dewitte, 2007). This removes gender as a potential source of variance. Participants were primarily Caucasian (79.7%), with some African-American (8.6%), Asian (4.6%), and Latino (3.3%), and 3.8% identified as another race. On average, participants were 19.0 (SD = 1.7, range 18-31) years old.

The semester ended before the last four experimental sessions could be conducted. Many subjects had to be excluded: 41 subjects were excluded because the RA marked the session as having some error, 3 were excluded for gameplay data that indicated an error of game assignment (e.g., dying in the easy game), 13 were excluded for missing data on the primary DV, and 114 were excluded for indicating awareness of the hypothesis. After exclusions, the final effective sample size was 275.

**Disclosures**

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. Hypotheses and sample size were preregistered at https://osf.io/cwenz/?view\_only=17755261098c4fc19a643572beb77855. All measures, materials, data, and analytic code are also available at that URL.

**Measures**

**2D:4D ratio.** Participants placed their hands on a flatbed scanner, fingers held together and fully extended. The distance from tip to basal crease of each index and ring finger was measured using the caliper tool in the GNU Image Manipulation Program (The GIMP Team, n.d.), a freeware Photoshop-like tool. 2D:4D ratios were created for each hand by taking the ratio of lengths of the index and ring fingers. Five coders provided measurements in this fashion, with each scan coded by at least two coders. Inter-rater reliability was assessed using a one-way, mixed, consistency, average-measures intra-class correlation (Hallgren, 2012; McGraw & Wong, 1996) using the psych package for R (Revelle, 2017). The resulting ICCs were excellent (ICC3k = .94 for left 2D:4D, .88 for right 2D:4D), indicating high agreement across coders and minimal loss of power due to measurement error.

**Cold pressor task.** Participants had an opportunity to aggress against their partner by assigning the partner to immerse his fist in a bucket of painfully-cold water for an amount of time (Pedersen, Bushman, Vasquez, & Miller, 2008). The cover story described the cold pressor as a form of distraction used in a subsequent task measuring decision-making under distraction. Participants learned that cold pressor immersion is unpleasant by sampling the cold water for five seconds before making their assignment. The participant then assigned the partner to a duration of cold-water immersion on a 9-point scale, ranging from 0 to 80 seconds in 10-second intervals. This measure can be quantified in only one way (i.e., 1-9 rating), eliminating concerns about flexible quantification methods associated with the competitive reaction time measure of aggression (Elson et al., 2014).

**Manipulation checks.** Participants completed a questionnaire assessing the efficacy of the various parts of the experimental manipulation. First, participants rated their partner’s feedback as pleasant or irritating (6 items). Then, participants rated the video game they played on a number of dimensions, including how violent, exciting, and challenging it was (18 items). All items were rated on a 1 (strongly disagree) to 7 (strongly agree) Likert scale. Participants then rated their degree of experience with video games, first-person shooter video games, and playing video games with a keyboard and mouse. Finally, participants provided demographic information about themselves.

**Probe for suspicion.** Research assistants attempted an oral funneled debriefing. Following this oral debriefing, participants completed a questionnaire intended to imitate a funneled debriefing. This debriefing questionnaire started with broad questions about the study and then grew increasingly specific, asking whether anything seemed strange about the study, the aggression measure, or the other participant in the study. Participants who indicated that the study was about the effects of video games on aggression without picking any lures (debriefing questionnaire, item 1) were marked as suspicious and excluded from analysis.

**Materials**

**Modified video games.** Four modified versions of the video game *Doom II* (iD Software, 1994) were created using software modification tools. These four versions were designed to create a 2 (Difficulty: Easy, Difficult) x 2 (Violence: Nonviolent, Violent) design.

Violent content was manipulated by changing the graphics, sounds, and story of the game while leaving the controls and enemy behavior constant. In the violent version, enemy graphics and sounds were borrowed from *Brutal Doom* (Abenante, 2012), a modified form of *Doom II* that makes the game more violent. In this game, defeated enemies exploded into fountains of gore and severed limbs. Participants in this condition were told to kill all the aliens. In the nonviolent version, enemy graphics and sounds were borrowed from *Chex Quest* (Digital Café, 1996), a modified version of *Doom II* that replaces the enemies with silly-looking booger aliens. Participants in this condition were told that the aliens were lost and confused and needed to be sent home with the “zorcher”, a tool resembling a remote controller.

The difficulty of the games was manipulated by changing the enemies’ artificial intelligence. In the difficult version of the game, the enemies fought back, using weapons in the violent game and throwing boogers in the nonviolent game. In the easy version of the game, enemies would not attack the player and instead walked slowly towards the player and waited to be killed or zorched.

The games were also programmed to track some statistics about the player’s performance. These gameplay variables included player deaths, player kills, wounds received, bullets fired, shotgun shells fired, and distance progressed. These were chiefly collected for quality control purposes to make sure that players assigned to the easy condition did not receive wounds and that all players fired their weapons / zorchers and killed / zorched some monsters.

**Procedure**

Participants arrived at the lab in pairs and were immediately escorted to separate adjacent rooms. In the case that only one participant was present, a male research assistant or graduate student would pretend to be the other participant. Following consent, their hands were scanned. Participants were able to see each other as scans were taken, demonstrating the presence of another participant in the study. After scanning, participants returned to their desks.

Participants were provoked by their partner in a procedure adapted from Bushman and Baumeister (1998). Participants were given an envelope, a sheet of loose-leaf paper, and a printed essay prompt. They were informed that the first task was to write a five-minute persuasive essay of their personal views on abortion, which would later be judged by the other participant. At the end of five minutes, the essays were collected so that they purportedly could be exchanged with the other participant. All participants wrote an essay that established an unambiguous view on abortion.

During the exchange, each participant received a fake, premade essay designed to oppose their beliefs (pro-life for pro-choice and vice-versa). Participants rated the essay, then put the essay and evaluation in the partner’s envelope, which was then taken from the room, ostensibly for data entry.

Participants then played their assigned version of the video game for 15 minutes. Each received a cover story that explained the story and controls of the game.

The cold pressor consisted of a pitcher of water kept in the laboratory refrigerator. Five minutes before the end of the gameplay session, the research assistant added a dozen ice cubes to the pitcher.[[1]](#footnote-2)

When the game session ended, the research assistant brought the cold pressor pitcher and a towel into the room, recorded the gameplay variables, and quit the game. The assistant then prepared to open an E-Prime script ostensibly containing the next task. Participants were told that the next portion of the experiment involved performing a computer task while distracted by cold water (the cold pressor). Participants sampled the cold pressor for five seconds.

The research assistant then provoked the participant by bringing the participant’s original envelope into the room and showing him the partner’s rating of his essay. The partner had rated all dimensions as between -8 and -10 in quality and commented, “This is the stupidest thing I’ve ever read.”[[2]](#footnote-3) The participant was then asked to assign their partner to an amount of distraction (that is, cold pressor exposure). The researcher explained that, to avoid experimenter bias, participants were being asked to randomly assign each other to the various levels of distraction.

Finally, participants were told that the experiment was running out of time and that the distraction task would be skipped. Participants completed post-questionnaires asking them to rate the games, their partner’s feedback, and what they suspected was the purpose of the study. Participants were then fully debriefed and dismissed.

**Results**

**Manipulation Check**

**Game manipulation.** Participant ratings on the post-questionnaires were submitted to 2 (Violence) 2 (Difficulty) ANOVAs. The manipulation was highly effective: participants indicated that the violent game (*M* = 5.3, *SD* = 1.6) was much more violent than the nonviolent game (*M* = 2.2, *SD* = 1.3; *d* = 2.1 [1.8, 2.4]). A latent factor was extracted to represent experienced challenge. Participants rated the difficult game as more challenging than the easy game, *t*(243) = 6.15, *p* < .001, *d* = 0.74 [0.50, 0.99]. The difficult game was not seen as more violent, nor was the violent game seen as more difficult.

Players generally did not perceive themselves to have behaved aggressively during the game (*M* = 3.6, *SD* = 1.78 in the violent game, *M* = 2.76, *SD* = 1.55 in the less-violent game, where 4 represents “Neither agree nor disagree”). Ratings of one’s own aggressive behavior were significantly influenced by the game’s violent content (*t*(245) = 3.96, *p* < .001, *d* = 0.48 [0.24, 0.72]), but not by the game’s difficulty (*t*(245) = -0.49, *p* = .627, *d* = -0.06 [-0.30, 0.18]). An interaction was observed such that the difference between violent and non-violent games was larger when the game was easy, *t*(245) = -2.09, *p* = .038, *d* = -0.26 [-0.51, -0.01].

**Provocation.** Mean evaluations of the participants’ interactions with the partner were also assessed. Participants generally indicated that they were irritated (*M* = 5.0, *SD* = 1.7), angered (*M* = 4.2, *SD* = 1.8), and annoyed (*M* = 4.9, *SD* = 1.8) by their partner. Furthermore, they were neither happy (*M* = 2.4, *SD* = 1.4) nor pleased (*M* = 2.2, *SD* = 1.4) with their partner and found the feedback unhelpful (*M* = 1.7, *SD* = 1.3).

To determine whether the cold pressor dependent variable was a sensitive measure of aggression, we tested whether participants more provoked by the feedback gave higher cold pressor assignments. Parallel analysis suggested a two-factor solution for participants’ ratings of their interaction with their partner. Factors were extracted using oblimin rotation. The first factor accounted for 52% of the variance and had loadings indicating provocation: .77, .76, and .67 for irritation, anger, and annoyance and -.25, .02, and .02 for happiness, helpfulness, and pleasure. This provocation factor was then used as a linear predictor of cold pressor assignment. The relationship was moderately strong, *t*(249) = 5.73, *p* < .001, *r* = .33 [.22, .43], suggesting that the cold pressor measure was indeed influenced by participants’ intent to aggress. A scatterplot and loess regression line are provided in the supplement.

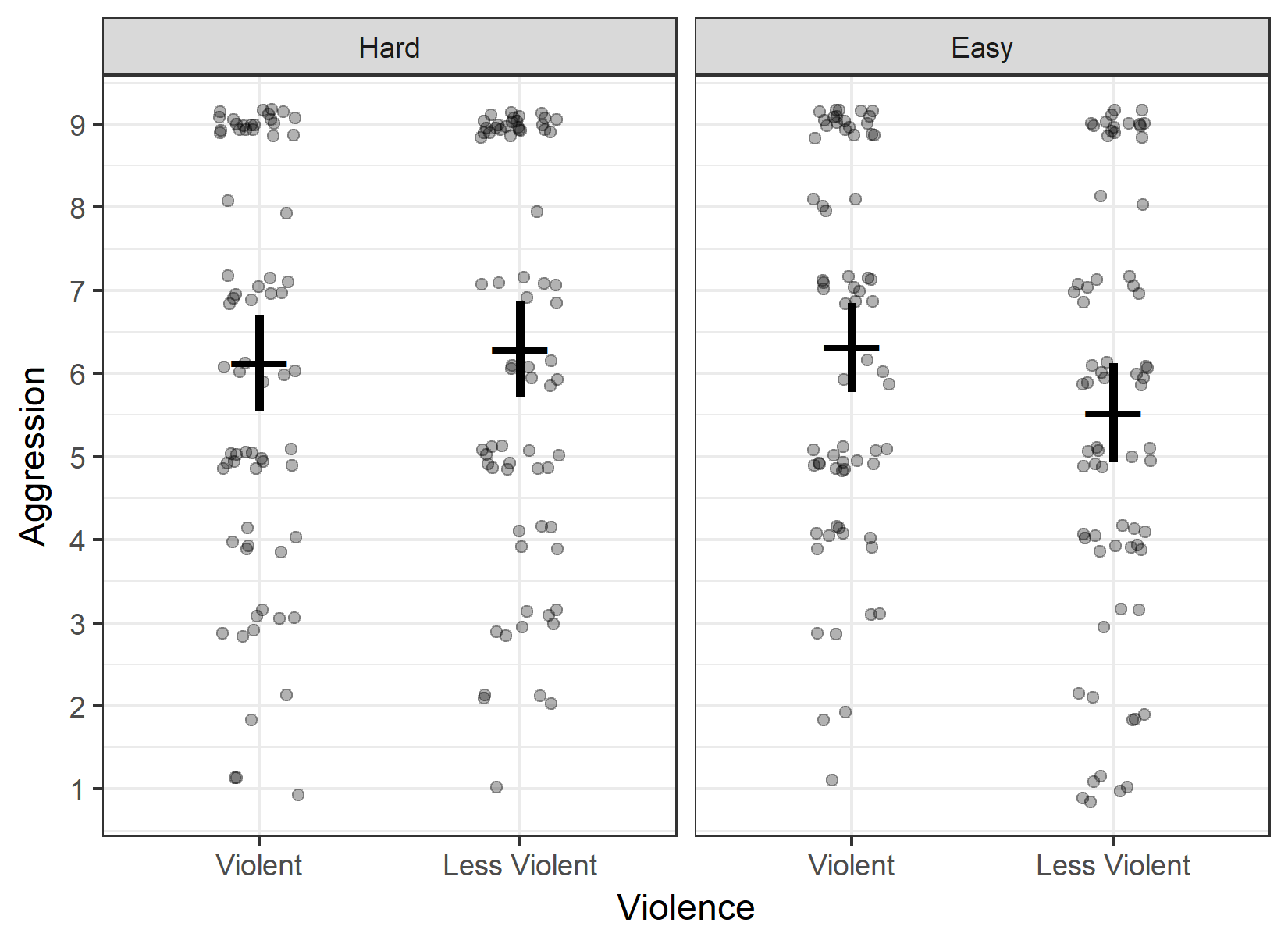
If violent games increase hostile appraisals, one might expect participants in the violent-game condition to rate their interaction more negatively. A 2 (Violence) 2 (Difficulty) ANOVA was conducted to determine whether the game played influenced participants’ ratings of the interaction. Effects were small and not statistically significant (violence, *t*(247) = -0.28, *p* = .777, *d* = -0.03 [-0.27, 0.20]; difficulty, *t*(247) = -0.17, *p* = .867, *d* = -0.02 [-0.26, 0.22]; Violence × Difficulty, *t*(247) = -0.86, *p* = .392, *d* = -0.10 [-0.34, 0.13]), suggesting that the game played had a minimal influence on participants’ reaction to the feedback.

**Conventional General Linear Models**

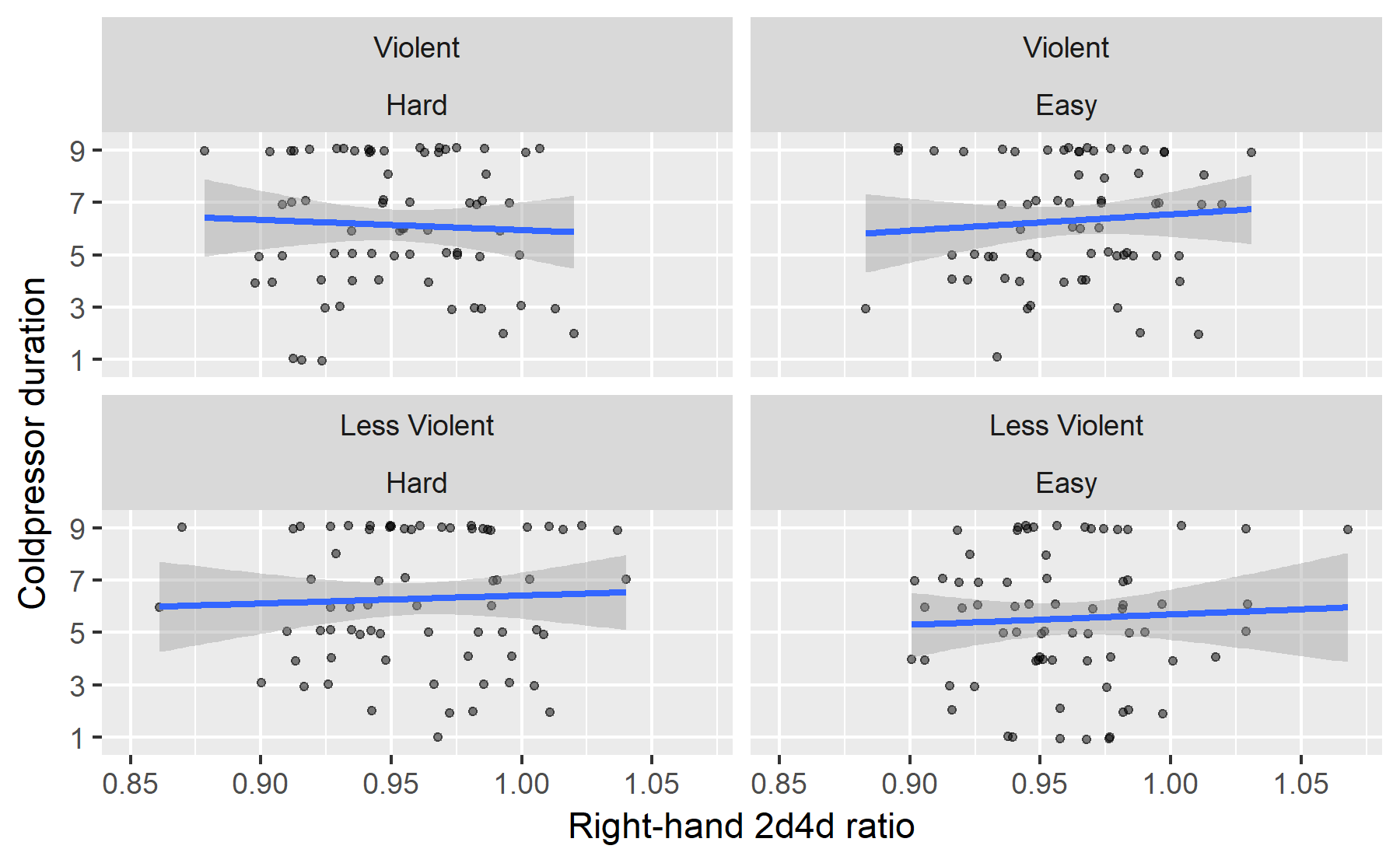
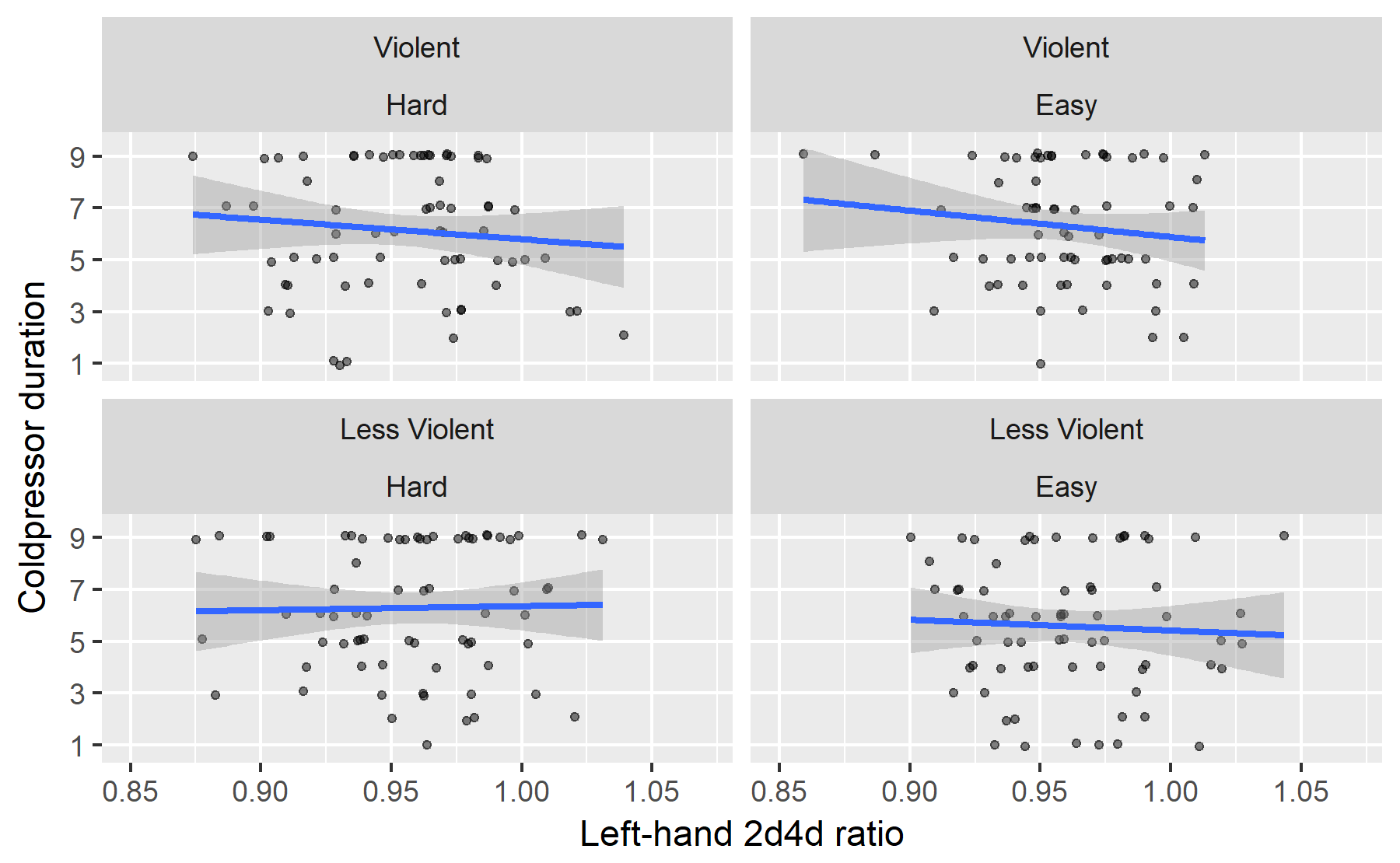
General linear models were used to look for main effects and interactions of game difficulty, game violence, and 2D:4D ratio. These tests were preregistered. Two models were used to look for effects of left and right 2D:4D ratio separately. Factors were contrast-coded and 2D:4D ratios were standardized to preserve orthogonality of parameter estimates. Cell means and SDs are provided in Table 1 and visualized in Figure 1.

Neither model found any significant effects. Neither game violence (*t*(265) = 0.90, *p* = .371, *d* = 0.11 [-0.13, 0.35]), game difficulty (*t*(265) = 0.85, *p* = .395, *d* = 0.10 [-0.13, 0.34]), nor their interaction (*t*(265) = -1.52, *p* = .129, *d* = -0.18 [-0.42, 0.05]) significantly predicted aggression. Additionally, neither left-hand 2D:4D (*t*(265) = -1.11, *p* = .266, *r* = -.07 [-.18, .05]) nor right-hand 2D:4D (*t*(266) = 0.52, *p* = .602, *r* = .03 [-.09, .15]) had a significant main effect on aggressive behavior (Figure 2). No higher-order interactions involving 2D:4D ratio of either hand were statistically significant. Full model output is summarized in Tables 2 and 3.

Figure 1. Mean aggression across conditions



Cold pressor assignments across the 2 (Violence: Violent (Brutal Doom), Less Violent (Chex Quest)) × 2 (Difficulty: Hard, Easy) design. Individual subjects are represented as translucent points. Slight vertical and horizontal jitter has been added to reduce overplotting. Means and CIs are overlaid as the horizontal cross and vertical whisker.

Figure 2. Null relationship between 2D:4D and aggression  

Scatterplots illustrating the relationship between 2D:4D and aggression in each condition. Slight vertical jitter has been added to reduce overplotting. Relationships are consistently near zero.

The earlier manipulation and sensitivity check indicated that much of the variance in aggression could be predicted by experienced provocation. Because this provocation was generally independent of the experimental condition, it was appropriate to try it as a covariate to increase statistical power. However, adding provocation as a covariate did not reveal significant effects. The effect of violence was *t*(246) = 0.78, *p* = .434, *d* = 0.09 [-0.14, 0.33], the effect of difficulty was *t*(246) = 1.08, *p* = .283, *d* = 0.13 [-0.11, 0.37], and their interaction was *t*(246) = -1.00, *p* = .318, *d* = -0.12 [-0.36, 0.12]. Effects of left-hand and right-hand 2D:4D remained nonsignificant (*t*(246) = -1.86, *p* = .065, *r* = -.12 [-.24, .01] and *t*(248) = -0.31, *p* = .755, *r* = -.02 [-.14, .11], respectively).

**Bayesian ANOVA**

Models were compared using the BayesFactor package for R (Morey & Rouder, 2014). The scale of the effect size under the alternative hypothesis was specified as *δ* ~ Cauchy(.4), consistent with the effect size reported in meta-analysis (Anderson et al., 2010). Models were generated to represent all possible combinations of main effects and/or interactions. Models including interactions were required to include all lower-order interactions and main effects. All models were compared to a null-hypothesis model of no effects: *δ* = 0.

Of all the models, the null-hypothesis model was best supported by the data. Models of main effects of Violence, Difficulty, left-hand 2D:4D, or right-hand 2D:4D were each outperformed by the null model (Bayes factors = 3.61, 3.81, 4.40, and 6.53 in favor of the null, respectively). Higher-order interactions were not supported by the data, either. Evidence was ambiguous regarding a Violence × Difficulty interaction (BF = 1.42 favoring the null). Neither violence nor difficulty interacted with 2D:4D of the left hand (BF01 = 3.97, 4.84, respectively) or 2D:4D of the right hand (BF01 = 4.97, 4.68). The Violence × Difficulty × 2D:4D interaction was not supported (left-hand BF01 = 3.59, right-hand BF01 = 3.16).

Experienced provocation was added to the model as a predictor. An effect of provocation was strongly supported by the evidence (BF10 = 1.04×106). However, addition of this covariate did not improve the strength of evidence for main effects of violence (BF01 = 5.04), difficulty (BF01 = 3.62), or 2D:4D (left hand, BF01 = 1.26; right hand, BF01 = 6.13). Taken together, these results indicate that aggression could be predicted by experienced provocation but not by game condition.

**Non-local Bayesian prior.** In the Bayesian hypothesis tests provided above, we use a non-directional, non-specific alternative hypothesis scaled roughly to the magnitude of the expected effect. While this is a useful hypothesis to test, it would also be useful to compare the obtained results against a more specific alternative hypothesis representing the effect as estimated from previous meta-analysis, δ = .43 (.35, .52) (Anderson et al., 2010).

The main effect of Violence was *d* = 0.11 [-0.13, 0.35]. An online Bayes factor calculator (Dienes, 2008) was used to compare the evidence for H0: *δ* = 0 relative to H1: *δ* = .43 [.35, .52]. The obtained Bayes factor substantially preferred the null, BF01 = 14.2.

Proponents have suggested that the Anderson et al. (2010) estimate may be an overestimate due to publication bias, but that after adjustment for publication bias the effect is still approximately *d* = .30. The Bayes factor calculator was used to compare the evidence for H0: *δ* = 0 relative to H2: *δ* = .30 [.20, .40]. The obtained Bayes factor still preferred the null, but less so relative to this more modest estimate, BF02 = 2.0.

**Supplementary analyses**

Cold pressor assignments were found to be non-normally distributed. To address this non-normality, the data were also analyzed using censored regression, logistic regression, and ordinal regression. Censored regression was used to attempt to model responses greater than 9, and logistic regression was used to model the probability of a 9 response vs. all other responses. These methods did not yield substantively different conclusions (i.e., no parameters were significant). See the supplement for details.

**Exploratory analyses**

Exploratory analyses examined other potential correlates of aggression. In general, we did not observe correlations between participants’ self-reported history of game use or participants’ in-game behavior. See the supplement for details.

**Discussion**

Results indicate that when game stimuli are carefully controlled, the effects of fifteen minutes of violent and/or difficult gameplay on aggressive behavior may be small and indistinguishable from zero. This suggests that the effects of brief violent video game play on aggressive outcomes may be smaller and less robust than the published research literature would indicate (see also Engelhardt, Mazurek, Hilgard, Rouder, & Bartholow, 2015; Hilgard, Engelhardt, & Rouder, 2017; McCarthy, Coley, Wagner, Zengel, & Basham, 2016). Researchers may need to reevaluate whether violent game manipulations are useful for revealing the causes and mechanisms of aggression. Further research will also be necessary to determine whether, and under which conditions, competitive or frustrating gameplay causes aggression (Adachi & Willoughby, 2011; Przybylski et al., 2014).

2D:4D digit ratio also failed to predict aggressive behavior among participants. The current results cast doubt on 2D:4D as an index of prenatal testosterone and a predictor of aggressive behavior (see also Hönekopp & Watson, 2011; Voracek, 2014). The sample size of the current research is considerably larger than many other studies reporting significant associations between 2D:4D ratio and aggression (e.g., Millet & Dewitte, 2007, 2009).

The presented manipulation and sensitivity checks suggest that the null results are not due to failures of the methodology. First, participants indicated that the violent game was much more violent than the nonviolent game. Second, participants were generally irritated with their essay feedback. These indicate that both the game manipulation and the essay provocation were effective. Third, the cold pressor measure of aggression was moderately sensitive (*r* = .33) to participants’ irritation with their partners. This sensitivity suggests that the null result is not due simply to the unusual distribution of the data or an overall invalidity of the cold pressor measure.

**Limitations**

Three factors may have reduced the effect size relative to previous research. First, it is possible that the nonviolent *Chex Quest* game involves sufficient violence to cause an increase in aggression, eliminating the difference between conditions. One study has claimed that the effect of cartoon E-rated violence is as strong as that of explicit M-rated violence (Anderson, Gentile, & Buckley, 2007). This seems unusual; compared to mild violent content, exposure to more extreme violent content should be more desensitizing, activate more aggressive thoughts, stimulate more aggressive feelings, and reward more aggressive behavior. Furthermore, our participants generally disagreed that *Chex Quest* involved violence or that their in-game behavior was aggressive. Still, it is possible that an effect was not found in the present study because *Chex Quest* causes some increase in aggression, reducing the effect size when compared against *Brutal Doom*. Future research may test the dose-response curve of violent content and aggressive behavior.

Second, because a plurality of participants (29%) gave the maximum possible aggressive response, it is possible that our measure is not sensitive to the influence of violent games. In the context of provocation, there may be a ceiling effect that compresses scores and reduces the sensitivity of the measure. On the other hand, only a minority of responses were at the maximum, and logistic binomial and ordinal regression observed no differences between groups in maximum responding. Furthermore, there was sufficient variability in both provocation and aggression that we were able to observe a moderate (r = .33) correlation between the two, suggesting that the outcome remained a valid and sensitive measure of aggression. Since all participants were provoked, provocation is not a source of noise variance; at most, effect sizes might be reduced by restriction of range. Finally, we note that other studies have found violent-game effects on aggression in the context of provocation (for example, Carnagey & Anderson, 2005 use the same essay provocation and report significant effects on a noise blast task). Nevertheless, restriction of range in the outcome could reduce the effect size. We encourage researchers to report the properties and test the validity of measures of aggressive behavior.

Finally, many participants indicated awareness of the research hypothesis and were discarded. This may have been due, in part, to the redundant process of oral funneled debriefing and questionnaire funneled debriefing, which may have increased awareness of the hypothesis following collection of the primary outcome. We chose to be conservative in our quality checks so as not to overstate the evidence for the null hypothesis. Nevertheless, one might be concerned that still more participants were hypothesis-aware, reducing the observed effect size through reduction of internal validity or through reactance (Bender, Rothmund, & Gollwitzer, 2013). Researchers may find value in establishing standardized practices in deception and debriefing.

**Summary**

We found evidence that brief exposure to violent games does not cause aggressive behavior. It is uncertain whether laboratory paradigms involving brief exposure to violent video games can reveal the causes of aggression. 2D:4D similarly does little to predict aggression in a laboratory experiment. One might question the validity of 2D:4D as an index of prenatal testosterone or whether prenatal testosterone predicts aggression. Research may benefit from addressing sources of irreplicabilty in this literature and considering other potential causes of aggression.

References

Abenante, M. (2012). Brutal Doom. Retrieved from http://www.moddb.com/mods/brutal-doom

Adachi, P. J. C., & Willoughby, T. (2011). The effect of video game competition and violence on aggressive behavior: Which characteristic has the greatest influence? *Psychology of Violence*, *1*(4), 259–274. https://doi.org/10.1037/a0024908

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

Anderson, C. A., & Carnagey, N. L. (2009). Causal effects of violent sports video games on aggression: Is it competitiveness or violent content? *Journal of Experimental Social Psychology*, *45*(4), 731–739. https://doi.org/10.1016/J.JESP.2009.04.019

Anderson, C. A., & Dill, K. E. (2000). Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life. *Journal of Personality and Social Psychology*, *78*(4), 772–790. https://doi.org/10.1037//O022-3514.78.4.772

Anderson, C. A., Gentile, D. A., & Buckley, K. E. (2007). *Violent video game effects on children and adolescents : theory, research, and public policy*. Oxford University Press.

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., … Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, *136*(2), 151–173. https://doi.org/10.1037/a0018251

Bender, J., Rothmund, T., & Gollwitzer, M. (2013). Biased Estimation of Violent Video Game Effects on Aggression: Contributing Factors and Boundary Conditions. *Societies*, *3*(4), 383–398. https://doi.org/10.3390/soc3040383

Bushman, B. J., & Anderson, C. A. (2002). Violent Video Games and Hostile Expectations: A Test of the General Aggression Model. *Personality and Social Psychology Bulletin*, *28*(12), 1679–1686. https://doi.org/10.1177/014616702237649

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*(1), 219–229. https://doi.org/10.1037/0022-3514.75.1.219

Carnagey, N. L., & Anderson, C. A. (2005). The Effects of Reward and Punishment in Violent Video Games on Aggressive Affect, Cognition, and Behavior. *Psychological Science*, *16*(11), 882–889. https://doi.org/10.1111/j.1467-9280.2005.01632.x

Carré, J. M., McCormick, C. M., & Hariri, A. R. (2011). The social neuroendocrinology of human aggression. *Psychoneuroendocrinology*, *36*(7), 935–944. https://doi.org/10.1016/j.psyneuen.2011.02.001

Cohen-Bendahan, C. C. C., Buitelaar, J. K., van Goozen, S. H. M., Orlebeke, J. F., & Cohen-Kettenis, P. T. (2005). Is there an effect of prenatal testosterone on aggression and other behavioral traits? A study comparing same-sex and opposite-sex twin girls. *Hormones and Behavior*, *47*(2), 230–237. https://doi.org/10.1016/J.YHBEH.2004.10.006

Dienes, Z. (2008). *Understanding psychology as a science : an introduction to scientific and statistical inference*. Palgrave Macmillan.

Digital Café. (1996). Chex Quest. Retrieved from http://www.chexquest.org/index.php?action=downloads;cat=1

Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (2015). Comparing apples and oranges? Evidence for pace of action as a confound in research on digital games and aggression. *Psychology of Popular Media Culture*, *4*(2), 112–125. https://doi.org/10.1037/ppm0000010

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the competitive reaction time task in aggression research. *Psychological Assessment*, *26*(2), 419–432. https://doi.org/10.1037/a0035569

Elson, M., & Quandt, T. (2016). Digital Games in Laboratory Experiments: Controlling a Complex Stimulus Through Modding. *Psychology of Popular Media Culture*, *5*(1), 52–65. https://doi.org/10.1037/ppm0000033

Engelhardt, C. R., Mazurek, M. O., Hilgard, J., Rouder, J. N., & Bartholow, B. D. (2015). Effects of Violent-Video-Game Exposure on Aggressive Behavior, Aggressive-Thought Accessibility, and Aggressive Affect Among Adults With and Without Autism Spectrum Disorder. *Psychological Science*, *26*(8). https://doi.org/10.1177/0956797615583038

Ferguson, C. J., & Kilburn, J. (2010). Much ado about nothing: The misestimation and overinterpretation of violent video game effects in Eastern and Western nations: Comment on Anderson et al. (2010). *Psychological Bulletin*, *136*(2), 174–178. https://doi.org/10.1037/a0018566

Greitemeyer, T., & Mügge, D. O. (2014). Video Games Do Affect Social Outcomes. *Personality and Social Psychology Bulletin*, *40*(5), 578–589. https://doi.org/10.1177/0146167213520459

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22833776

Hilgard, J., Engelhardt, C. R., Bartholow, B. D., & Rouder, J. N. (2017). How much evidence is p &gt;.05? Stimulus pre-testing and null primary outcomes in violent video games research. *Psychology of Popular Media Culture*, *6*(4). https://doi.org/10.1037/ppm0000102

Hilgard, J., Engelhardt, C. R., & Rouder, J. N. (2017). Overstated evidence for short-term effects of violent games on affect and behavior: A reanalysis of Anderson et al. (2010). *Psychological Bulletin*, *143*(7), 757–774. https://doi.org/10.1037/bul0000074

Hönekopp, J., & Watson, S. (2011). Meta-analysis of the relationship between digit-ratio 2D:4D and aggression. *Personality and Individual Differences*, *51*(4), 381–386. https://doi.org/10.1016/J.PAID.2010.05.003

iD Software. (1994). Doom II. Rockville, MD: ZeniMax Media.

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., & Manning, J. T. (2004). 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Human Development*, *77*(1–2), 23–28. https://doi.org/10.1016/J.EARLHUMDEV.2003.12.002

Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Human Reproduction*, *13*(11), 3000–3004. https://doi.org/10.1093/humrep/13.11.3000

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

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

McIntyre, M. H., Barrett, E. S., McDermott, R., Johnson, D. D. P., Cowden, J., & Rosen, S. P. (2007). Finger length ratio (2D:4D) and sex differences in aggression during a simulated war game. *Personality and Individual Differences*, *42*(4), 755–764. https://doi.org/10.1016/J.PAID.2006.08.009

Millet, K. (2011). An interactionist perspective on the relation between 2D: 4D and behavior: An overview of (moderated) relationships between 2D: 4D and economic decision making. *Personality and Individual Differences*, *51*(4), 397–401. Retrieved from http://www.sciencedirect.com/science/article/pii/S0191886910001996

Millet, K., & Dewitte, S. (2007). Digit ratio (2D:4D) moderates the impact of an aggressive music video on aggression. *Personality and Individual Differences*, *43*(2), 289–294. https://doi.org/10.1016/J.PAID.2006.11.024

Millet, K., & Dewitte, S. (2009). The presence of aggression cues inverts the relation between digit ratio (2D:4D) and prosocial behaviour in a dictator game. *British Journal of Psychology*, *100*(1), 151–162. https://doi.org/10.1348/000712608X324359

Pedersen, W. C., Bushman, B. J., Vasquez, E. A., & Miller, N. (2008). Kicking the (Barking) Dog Effect: The Moderating Role of Target Attributes on Triggered Displaced Aggression. *Personality and Social Psychology Bulletin*, *34*(10), 1382–1395. https://doi.org/10.1177/0146167208321268

Pedersen, W. C., Vasquez, E. A., Bartholow, B. D., Grosvenor, M., & Truong, A. (2014). Are You Insulting Me? Exposure to Alcohol Primes Increases Aggression Following Ambiguous Provocation. *Personality and Social Psychology Bulletin*, *40*(8), 1037–1049. https://doi.org/10.1177/0146167214534993

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. https://doi.org/10.1037/a0034820

Revelle, W. (2017). psych: Procedures for Personality and Psychological Research. Evanston, Illinois: Northwestern University.

Slotter, E. B., & Finkel, E. J. (2011). I3 Theory: Instigating, impelling, and inhibiting factors in aggression. In P. R. Shaver & M. Mikulincer (Eds.), *Human aggression and violence: Causes, manifestations, and consequences*. American Psychological Association.

The GIMP Team. (n.d.). GNU Image Manipulation Program. Retrieved from www.gimp.org

Voracek, M. (2014). No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D:4D): a comprehensive meta-analysis and critical evaluation of research. *Evolution and Human Behavior*, *35*(5), 430–437. https://doi.org/10.1016/J.EVOLHUMBEHAV.2014.05.009

Table 1. *Cell means of cold pressor assignment per condition.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Violence | Difficulty | *n* | *M* | *SD* |
| Brutal Doom | Hard | 69 | 6.13 | 2.46 |
| Brutal Doom | Easy | 67 | 6.31 | 2.24 |
| Chex Quest | Hard | 69 | 6.29 | 2.47 |
| Chex Quest | Easy | 70 | 5.53 | 2.54 |

Table 2. *Effects of condition and left 2D:4D on aggression*

|  |  |  |  |
| --- | --- | --- | --- |
| Predictor | *b* | *t* | *p* |
| Violence | 0.13 | 0.90 | 0.371 |
| Difficulty | 0.13 | 0.85 | 0.395 |
| Left 2D:4D | -0.17 | -1.11 | 0.266 |
| Vio × Diff | -0.22 | -1.52 | 0.129 |
| Vio × Left 2D:4D | -0.12 | -0.83 | 0.406 |
| Diff × Left 2D:4D | 0.07 | 0.46 | 0.646 |
| Vio × Diff × Left 2D:4D | -0.02 | -0.17 | 0.869 |

*Note:* All model terms have standard error 0.15 and 265 *df*.

Table 3. *Effects of condition and right 2D:4D on aggression*

|  |  |  |  |
| --- | --- | --- | --- |
| Predictor | *b* | *t* | *p* |
| Violence | 0.15 | 0.97 | .332 |
| Difficulty | 0.14 | 0.97 | .333 |
| Right 2D:4D | 0.08 | 0.52 | .602 |
| Vio × Diff | -0.24 | -1.59 | .113 |
| Vio × Right 2D:4D | -0.04 | -0.26 | .793 |
| Diff × Right 2D:4D | -0.09 | -0.62 | .537 |
| Vio × Diff × Right 2D:4D | -0.08 | -0.51 | .608 |

*Note:* All model terms have standard error 0.15 and 266 *df*.

1. We did not take measurements of the pitcher’s temperature at the time. However, recreating the cooling procedure at home, the first author observes that pitchers were likely cooled to 41-44°F (5-7°C). This is comparable to temperatures reported in similar work (e.g., Pedersen, Vasquez, Bartholow, Grosvenor, & Truong, 2014). [↑](#footnote-ref-2)
2. Originally, the comment read, “This is one of the worst essays I have ever read!” consistent with previous research. Participants generally found this to be suspicious and unbelievable, so we changed it to a more flippant and more credible insult. [↑](#footnote-ref-3)