Journal of Experimental Social Psychology 95 (2021) 104119

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Case Report

Violent and non-violent virtual reality video games: Influences on affect, aggressive cognition, and aggressive behavior. Two pre-registered experiments

Aaron Drummond [a,*], James D. Sauer [b], Christopher J. Ferguson [c], Peter R. Cannon [a], Lauren C. Hall [a]

[a] School of Psychology, Massey University, Palmerston North 4424, Manawatu, New Zealand
[b] School of Psychological Sciences, University of Tasmania, Australia
[c] Psychology, Stetson University, Florida, USA

ABSTRACT

Immersive Ambulatory Virtual Reality (IA-VR) video games are relatively new and highly immersive. Given speculation that immersion may increase psychological effects of playing games, we examined whether violent IA-VR (cf. flat-screen) games increase aggression. Here, we report the first experimental studies to assess the effects of violent and non-violent IA-VR (cf. flat-screen) games on affect, aggressive cognition, and behavior. In Study 1, 200 participants played violent or non-violent IA-VR or flat-screen games in a pre-registered protocol. IA-VR was associated with slightly higher positive affect, but no higher aggression than comparable flat-screen games. Although violent games (IA-VR and flat-screen) increased aggressive cognition, this did not translate to hostile affect or aggressive behavior. In Study 2, 96 participants played a violent IA-VR or flat-screen video game. Again, no effects of IA-VR were observed on aggressive cognition, behavior, or hostile affect. In both studies, the relationship between aggressive cognitions, behavior and hostile affect was virtually nil. Though further replications are required with a greater variety of stimulus games, our studies provide early evidence against the notion that IA-VR increase aggression compared to flat-screen games. The lack of relationship between aggressive cognition and behavior suggests potential weaknesses in fundamental assumptions of the General Aggression Model.

Video games are played by an estimated 2.8 billion people worldwide (Entertainment Software Association, 2017). Recently, Virtual Reality game systems have become cheaper and more accessible. These systems allow gamers to be visually and aurally isolated within, and physically explore, virtual environments. Some systems (e.g., HTC Vive/ Occulus Rift) allow users to walk within a predefined user space (minimum area = 9 m^2), tracking their ambulation within this region, further increasing immersion. We hereafter refer to these systems as Immersive Ambulatory Virtual Reality (IA-VR) systems to distinguish them from systems which only allow the user to remain seated in one position in a Virtual Reality environment.

Concerns persist that exposure to simulated violence might increase user aggression (Anderson et al., 2017; Prescott, Sargent, & Hull, 2018). The General Aggression Model (GAM) views aggression as primarily cognitive, proposing that consuming violent media causes rehearsal of aggressive cognitive schema. Activating and rehearsing aggressive schemata purportedly increases aggressive behavior in non-gaming contexts (Anderson & Carnagey, 2004; DeWall, Anderson, & Bushman, 2011). However, there is growing evidence that the relationship is more complex. Narrative context, player avatar, frustration and competitiveness moderate or negate violent content effects (Adachi & Willoughby, 2011; Przybylski, Deci, Rigby, & Ryan, 2014; Sauer, Drummond, & Nova, 2015). Here we tested the effects of an oft-speculated but under-researched moderator: Immersion. Specifically, do effects of exposure to violent game content vary depending on whether the gaming takes place in a standard or IA-VR context?

Meta-analyses suggest that violent games explain around 1% or less of variance in aggression (Hilgard, Engelhardt, & Rouder, 2017; Mathur and VanderWeele, 2019; Prescott et al., 2018). However, researchers disagree about whether effects of this size are meaningful (Ferguson & Kilburn, 2010; Funder & Ozer, 2019; Sauer & Drummond, 2020). Some meta-analyses even show null effects of violent content for the highest quality studies (Drummond, Sauer, & Ferguson, 2020). Nonetheless, it is important to understand factors that may amplify effects (Drummond, Sauer, & Garea, 2018). IA-VR games approximate violent behavior more closely than traditional games, for instance, with users physically aiming and firing virtual firearms. Further, feeling more present within games may increase aggressive outcomes for highly aggressive users (Farrar &

* Corresponding author.
E-mail address: a.drummond@massey.ac.nz (A. Drummond).

https://doi.org/10.1016/j.jesp.2021.104119
Received 10 June 2019; Received in revised form 11 February 2021; Accepted 14 February 2021

Krcmar, 2006; Krcmar & Farrar, 2009). Moreover, 3D-headset immersion (cf. flat-screen) may increase hostile affect post-game (Lull & Bushman, 2016). However, no research to date has examined the effects of IA-VR on aggression. Though early VR research found no effect on aggression (Arriaga, Esteves, Carneiro, & Monteiro, 2008; Tamborini, Eastin, Skalski, & Lachlan, 2004), the improved audio-visual quality and ambulatory controls in current-generation IA-VR may further heighten immersion. Indeed, recent research has shown that stressful experiences in the latest IA-VR equipment can produce substantial physiological changes in participants, including increased heart rate and skin conductivity, especially in response to fear-inducing and stressful virtual stimuli (Lin, 2017; Martens et al., 2019; Peterson, Furuichi, & Ferris, 2018). The HTC-Vive is realistic enough that is has been shown to be an effective learning tool for reducing fear of heights among acrophobic subjects (Freeman et al., 2018). These findings make it theoretically plausible that IA-VR may also increase aggressive affect, cognitions or behavior in response to violent content. IA-VR's newfound availability also makes the present investigation into their effects particularly timely and important from an applied perspective, as this is the first time that such technology is widely available to the public.

We explored whether IA-VR games increase aggression following violent gameplay. Participants played either a violent or non-violent game in flat-screen or IA-VR. Immersive Ambulatory VR may enhance aggressive schemata activation (Farrar & Krcmar, 2006). Thus, based on the GAM, we predicted that the effects of in-game violence on aggressive cognition, hostile affect and aggressive behavior would be greater for the IA-VR (cf. flat-screen) gaming conditions.

Study 1.

As a first step toward establishing whether there was any effect of IA-VR on aggressive thoughts, feelings or behaviors, we undertook a study investigating differences in aggressive behavior between 4 different games: a non-violent flat-screen game (Portal), a non-violent IA-VR game (Portal Stories), a violent flat-screen game (CoD: Infinite Warfare, Zombies mode) and a violent IA-VR game (Arizona Sunshine). Due to a paucity of games with identical IA-VR and flat-screen modes at the time of the experiment, we selected games on IA-VR and flat-screen platforms to be as similar as possible. We acknowledge this introduces some uncontrolled variability into the manipulation, but to foreshadow, we address this issue in Experiment 2.

1. Method

Participants & Design.

In this study, we report all measures, manipulations and exclusions. Two hundred people (133 male) recruited from a university campus in New Zealand participated in the study. We undertook a pre-registered protocol with clear stopping rules (see below). We terminated data collection when we reached our pre-specified maximum number of participants as we did not trip the other stopping rules in our protocol. Our pre-registration including stopping rules can be accessed at: https://aspredicted.org/d9a3y.pdf. Participants were randomly allocated to a 2 (virtual reality: flat screen; HTC Vive) by 2 (violent content: violent; non-violent) between-subjects experimental design.

We were attempting to gather evidence for the presence or absence of a greater effect on aggression of violent games in virtual reality than violent games on flat-screens (cf. non-violent games in both mediums). As such, in our view, the most important analysis in the present study was the analysis of how much wasabi participants administered during the wasabi task. To this end, we pre-registered a sequential Bayesian analysis. As per our preregistration document, once a minimum of 30 participants in each cell had been recruited (Simmons et al., 2011), we ran a Bayesian analysis once each week at the end of data collection for the week on the interaction between IA-VR condition and violent content on the amount of wasabi administered by participants. We pre-registered that we would cease data collection when we had strong evidence (i.e., a BF10 of either 10 or 0.1) for the presence or absence of an interaction. Due to financial constraints we also pre-registered that we would terminate data collection if we reached our maximum of 200 participants, or if we reached the 17th of July 2018 without meeting either of these targets. We reached 200 participants prior to our stopping date or reaching a Bayes factor of either 10 or 0.1, and thus ceased data collection.

Participants were required to have normal or corrected to normal vision. We initially pre-registered that long sighted participants would not be allowed to participate because the HTC Vive VR headset sits close to the eyes. However, pilot testing conducted after the pre-registration and prior to data collection confirmed that most participants were able to wear glasses in the HTC Vive with no issues and so we relaxed this exclusion criteria. Participants with any symptoms of vertigo were excluded from participating in the study.

As pre-registered, participants who scored four or more on the Generalised Anxiety Disorder 2 item scale (GAD-2; $n = 9$) were excluded from participating in the study due to the horror content of the games. These participants played a non-violent flat-screen game. Their data were discarded because the participants were not randomly allocated to a condition. One further participant was replaced due to failing to follow experimenter instructions, and one participant was replaced due to a compromised laboratory environment during testing (a loud social gathering held next to the laboratory). As these participants had received a participant number prior to starting the experiment, data from these participants was replaced with data from a new participant with an R after their participant number to denote them as replacements.

Sensitivity Power Analysis. Sensitivity power analysis assuming an alpha level of 0.05 and a power of 80% revealed that our sample was sufficient to reliably detect moderately sized differences between the means of two of our groups at an effect size of $d = 0.50$, and moderately sized interaction effects, $f = 0.29$.

Materials and Measures.

Games. Ideally, to compare IA-VR to flat-screen games, we would employ identical games which had both an IA-VR and flat-screen version. However, due to a paucity of games with identical IA-VR and flat-screen modes which were appropriate for such an experiment, we selected games on IA-VR and flat-screen platforms to be as similar as possible. To this end, the non-violent games were selected to be puzzle first person games. Specifically, the non-violent flat screen game used was "Portal", and the non-violent VR game was "Portal Stories". Both violent games were selected to be wave-based zombie first person shooters, with each wave (group) of enemies becoming progressively more difficult. The violent flat-screen game was "Call of Duty: Infinite Warfare" played in zombie survival mode, while the violent VR game was "Arizona Sunshine", a zombie game played in survival mode.

Flat screen games were displayed on a 50-in. high-definition 1080p television, with participants sitting approximately 2.45 m from the monitor. This had a 29.2° viewing angle. Virtual reality games were displayed using an HTC Vive containing two 1080p displays, one for each eye, resulting in a 145° viewing angle. Participants who played in virtual reality had a 3 m × 4 m ambulatory area in which they were free to walk around the environment. Movement beyond this radius was controlled through teleportation mechanics within the games.

Initial Questionnaire. Prior to gameplay, participants completed a brief questionnaire about their demographic characteristics. Specifically, participants were asked for their age, sex (male female), "How often do you play video games (Never, hardly ever, Once a month, Once a Week, Most days)"and "How good do you consider yourself to be at first person shooter video games; 1, novice – 7, expert". Participants were also screened for anxiety conditions with the GAD-2 (Donker et al., 2011). Any participant with a GAD-2 score of 4 or higher was excluded from participating in the study. Although we considered it unlikely that participants would become extremely anxious during gameplay, due to the immersion of virtual reality, we decided to exclude participants whom appeared to be highly anxious to safeguard against potential episodes of anxiety caused by the immersive virtual environment.

Aggressive Cognition. Following gameplay, participants undertook a word completion task as a measure of aggressive cognition (Carnagey & Anderson, 2005). Participants were presented with a series of 46 words with some letters missing and asked to fill in the blanks. Some words were able to be completed with either an aggressive or non-aggressive word. Note that only the first 46 word-fragments of this task were used in the present experiment due to time constraints.

Aggressive Behavior. Following the post-game questionnaires, participants were asked to engage in a wasabi task. This task is an adaptation of Fischer, Kastenmüller, and Greitemeyer's (2010) hot sauce task (see Sauer et al., 2015, for full details). In this task, participants were told that wasabi is a very spicy condiment that is not liked by many people and shown a brief video of a person eating, and displaying a negative reaction to, a large quantity of wasabi. Next, participants were told that as part of a pilot study on people's experiences of taste that they should dispense a quantity of wasabi into a cup, and that another person will later be asked to (a) eat all of the wasabi in that cup and (b) answer some questions about their experiences. Participants were supplied a syringe (without needle) containing 10mls of wasabi paste and asked to dispense any amount into a plastic cup. Participants were instructed that if they wished they did not need to administer any wasabi during this task. The amount of wasabi dispensed indexed aggressive behavior, as the participant is inflicting an unpleasant experience upon an ostensibly real other. This task has previously been used to successfully detect aggressive behavior following video game play under some conditions (Sauer et al., 2015).

Affect. Prior to and following gameplay, participants completed the Positive and Negative Affect Scale X (PANAS-X; Watson & Clark, 1999) as a measure of affect. Specifically, the PANAS-X measures the extent to which participants currently feel a range of emotions. Our primary interest was in participants' scores on the hostility subscale of this measure. However, we also pre-registered analyses on the general positive and negative affect scales as secondary analyses.

Other measures. Participants also completed several questions about their experiences with the game. As the virtual reality games and the flat screen games will be played using different controllers, participants were asked about how difficult they found it to control the game on a seven point scale (1, not at all, 7 extremely difficult; Przybylski et al., 2014). Participants also rated how immersive they found the experience and how novel they found the experience on seven-point scales (1, not at all immersive or novel - 7, extremely immersive or novel). These variables were intended to be used as exploratory variables to further investigate any differences in aggression following video game play.

Procedure.

Upon arriving at the laboratory participants filled out the screening questionnaire. Participants who qualified for the study completed the initial PANAS-X measure. Participants then received instructions on how to control the game they had been assigned, and given a 5 min practice period to familiarise themselves with the controls. Participants then played the game they were assigned for a period of 15 min. Participants were informed that they could stop playing at any time, and for any reason. Participants in the non-VR conditions were told that they may place the controller on the ground and stop playing the game at any time. Participants in the VR system were instructed that if they felt they need to stop playing, they could close their eyes, crouch to the floor and remove the headset. Participants were monitored by the experimenter during their play sessions. 194 participants played for the full 15 min period, and all participants played for more than 10 min. The average time of play for all 200 participants was 14 min and 57 s (*SD* = 21 s).

Following gameplay participants once again completed the PANAS-X, and filled out the word-stem completion task and the rest of the self-report measures. Participants then undertook the wasabi task. Following completion of the Wasabi task, participants were presented with a mood-stabilisation video (a short video about animal friendships) to counteract any potentially negative effects of the study.

Participants finally underwent a funnelled debriefing procedure in which they were probed for suspicion about the wasabi task and asked what they thought the hypotheses of the study were. This involved two questions. First, participants were asked what they believed the purpose of the study was. Participants who suggested the study may relate to the impact of the video game on aggression were coded as having guessed the hypothesis. Participants who suggested the study may relate to any other aspect of the effect of the video gameplay on subsequent cognition, emotion or behavior were coded as suspicious. All other participants were coded as not suspicious. Participants were then asked what they believed the purpose of the Wasabi task was. Participants who expressed that they thought the wasabi task was a measure of aggression were coded as having guessed the Wasabi task. Participants who indicated that they believed the Wasabi task might be a measure of some form of negative behavior (e.g., reduced empathy) were coded as suspicious of the Wasabi task. These variables were then employed for the purposes of supplementary analyses as detailed in the preregistration protocol. Finally, participants received a $10 gift voucher as a thank you for participating in the study.

2. Results

The complete dataset is publicly available https://osf.io/t9yrc/). Data were analysed in accordance with the pre-registration protocol document. To reduce potential bias, initial analyses were conducted by an analyst blind to the coding of the conditions. All analyses were conducted using Bayesian methods. All Bayesian ANOVA's were conducted using the default settings in JASP 0.10.0.0. This version does not allow for the altering of priors for Bayesian ANOVA analyses. For all analyses we adopted Kass and Raftery's (1995) criteria for interpreting the strength of evidence for (or against) an effect. Bayes Factors (BF) of 1–3.2 were considered inconclusive evidence for an effect, while BFs of 0.31–1 were considered inconclusive evidence for the null. BFs of 3.2–10 were considered substantial evidence for an effect, while 0.31 to 0.10 were considered substantial evidence for the null. BFs of 10–100 were considered strong evidence for an effect and 0.10–0.01 as strong evidence for the null. Finally, BFs of more than 100 were considered decisive evidence for an effect and less than 0.01 as decisive evidence for the null. Once interpretations of the analyses were agreed by the first four authors blind to the condition, data were de-blinded by the fifth co-author (LCH).

Confirmatory Analyses.

Aggressive Cognition. We calculated both the total number of aggressive words reported and the total number of aggressive words reported expressed as proportion of the total number of word stems attempted. These approaches yielded similar conclusions. Data for the word stem completion task showed normal values for skewness and kurtosis for both total number of aggressive words reported (skewness = 0.33, kurtosis = −0.36) and proportion of aggressive words reported (skewness = 0.42, kurtosis = 0.10). We ran Bayesian ANOVAs on the number of aggressive words reported and the proportion of aggressive words reported with VR condition (flat-screen; IA-VR) and violent content (non-violent; violent) as factors. There was ambiguous evidence, slightly in favour of the null, for the effect of VR condition on the number of aggressive words reported, $BF_{10} = 0.96$ and the proportion of aggressive words reported, $BF_{10} = 0.91$. There was substantial evidence that including the interaction term in the model did not improve the fit of the model to the data, $BF_{10} = 0.30$. There was also inconclusive evidence for a lack of interaction between VR condition and violent content on proportion of aggressive words reported, $BF_{10} = 0.83$.

As shown in Fig. 1, there was decisive evidence for a main effect of violent content on the total number of aggressive words reported, $BF_{10} = 2643.84$. Participants who played a violent game reported a greater total number of aggressive words (M = 7.37, SD = 2.55) than participants who played a non-violent game (M = 5.76, SD = 2.35); this effect

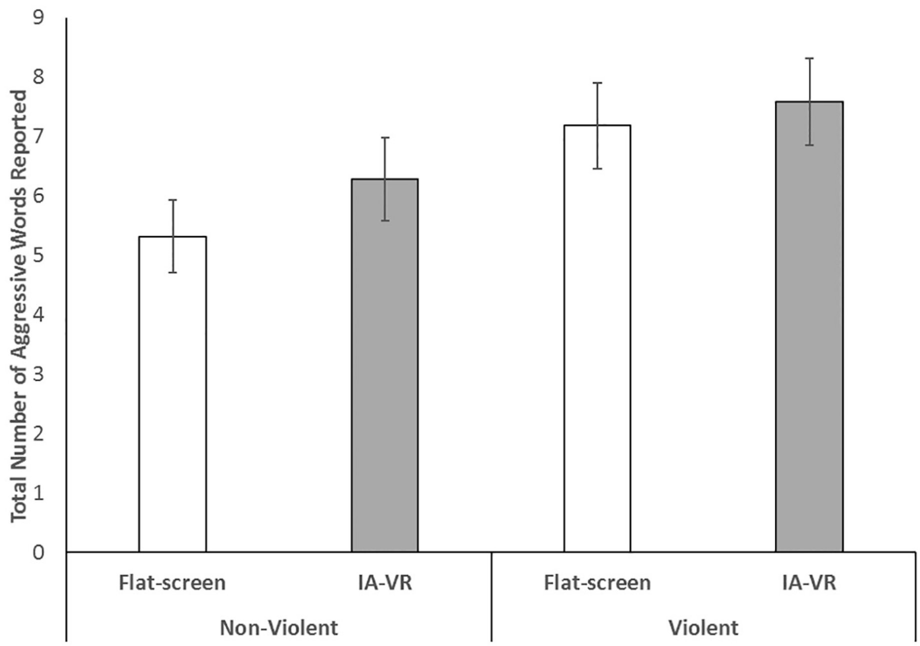

Fig. 1. Total number of aggressive words completed in the word stem completion task by experimental condition. Error bars represent 95% Credible Intervals. Participants who played a violent game reported more total aggressive words on average ($M = 7.37$, $SD = 2.55$) than participants who played a non-violent game ($M = 5.76$, $SD = 2.35$), a moderate sized effect, Cohen's $d = 0.66$.

was moderate in size, $d = 0.66$. The evidence was also decisive when proportion of aggressive words reported was used as the dependent measure, $BF_{10} = 17{,}884.91$. Participants who played a violent game reported a greater proportion of aggressive words on average ($M = 18.33\%$, $SD = 6.09\%$) than participants who played a non-violent game ($M = 14.17\%$, $SD = 5.40\%$); this effect was moderate in size, $d = 0.72$. Removing data from participants who guessed the hypothesis of the study yielded markedly similar results (Violence, $BF_{10} = 4984.58$, IA-VR $BF_{10} = 0.92$, Violence x IA-VR, $BF_{10} = 0.31$). Ten participants guessed the hypothesis of the study, and 63 reported suspicion of the hypothesis of the study. Twenty participants guessed the purpose of the Wasabi task while a further 63 participants were suspicious of the Wasabi task. Removing data from any participant who expressed suspicion of either the hypothesis of the study or the Wasabi task resulted in inconclusive evidence slightly favouring that IA-VR condition increased the total number of aggressive words reported for participants in VR ($M = 6.76$, $SD = 2.39$) compared to participants who played flat-screen games ($M = 5.61$, $SD = 2.26$), $BF_{10} = 2.57$. However, there was inconclusive evidence slightly favouring the null for the effect of VR condition on the proportion of aggressive words reported ($BF_{10} = 0.98$). Among this subset of participants there was inconclusive evidence for the null effect of violent content on the total number of aggressive words reported ($BF_{10} = 0.66$). There was substantial evidence that violent games increased the proportion of aggressive words reported ($M = 17.98$, $SD = 6.19$) compared to non-violent games ($M = 14.46$, $SD = 5.45$), $BF_{10} = 7.82$. These analyses yielded inconclusive evidence for the null interaction between violent content and IA-VR condition for the total number of aggressive words reported ($BF_{10} = 0.51$). However, there was strong evidence for an interaction between IA-VR condition and violence condition on the proportion of aggressive words reported ($BF_{10} = 14.63$). For flat-screen games, participants reported a greater proportion of violent words after playing a violent game ($M = 19.37$, $SD = 6.94$) than non-violent games ($M = 12.42$, $SD = 4.43$). However, for IA-VR games, participants reported a similar proportion of aggressive words following violent ($M = 16.82$, $SD = 5.39$) or non-violent ($M = 17.04$, $SD = 5.61$) game play.

Aggressive Behavior. Data for the wasabi task showed high levels of both skewness (2.12) and Kurtosis (4.14), but a square root transformation normalized the data (Skewness = 0.92, Kurtosis = 0.56). The conclusions drawn from analyses conducted with the untransformed and transformed data were identical. Thus, to aid interpretability and as specified in the preregistration document, we report the untransformed analyses here.

We undertook a Bayesian ANOVA with VR condition (flat-screen; IA-VR) and violent content (non-violent; violent) as factors. Bayesian analyses revealed evidence in favour of the null hypothesis. Bayes factors showed substantial evidence for null effects of the IA-VR, $BF_{10} = 0.20$, and violence manipulations, $BF_{10} = 0.21$, on aggressive behavior as measured by the amount of wasabi administered by participants. Evidence for the null model, compared to the model including the interaction and main effects of Violence condition and VR condition, was inconclusive $BF_{10} = 0.34$.

Removing data from participants who expressed suspicion that the wasabi task was a measure of aggression made little difference to the results, except to strengthen the evidence for a null interaction effect: There was now substantial evidence in favour of the null effect compared to an interaction, $BF_{10} = 0.29$. Evidence for the null effects of IA-VR Condition, $BF_{10} = 0.17$, and violence condition, $BF_{10} = 0.23$, on aggressive behavior remained consistent with analyses based on the full dataset. Similarly, removing data from participants who guessed the hypothesis of the study yielded similar results (Violence, $BF_{10} = 0.23$, IA-VR $BF_{10} = 0.18$, Violence x IA-VR, $BF_{10} = 0.31$). Removing all participants who expressed any suspicion of either the hypothesis or the Wasabi task resulted in equivalent evidence for and against the null hypothesis for the effects of the violent content (BF = 1.20). However, this analysis still showed substantial support for the null effect of IA-VR condition ($BF_{10} = 0.23$), and inconclusive evidence for the interaction between IA-VR and violence condition ($BF_{10} = 0.73$). No analyses showed evidence in favour of the alternative hypothesis.

Hostile Affect. Initial normality checks indicated that for the hostile affect subscale of the PANAS-X, both baseline (skewness = 1.82, kurtosis = 4.53) and post-game data (skewness = 1.71, kurtosis = 4.24) showed heavy-tailed kurtosis. Undertaking a square root transformation on the data reduced this kurtosis but did not reduce the kurtosis values to below 2 for either baseline (skewness = 1.40, kurtosis = 2.90) or post-game data (skewness = 1.22, kurtosis = 2.38). Thus, as specified in our preregistration document, we undertook a logarithmic transformation on the data. This reduced the kurtosis values to normal levels for both baseline (skewness = 0.98, kurtosis = 1.75) and post-game data (skewness = 0.76, kurtosis = 1.18). Most conclusions were similar when the untransformed data were used, although for one interaction the untransformed data showed inconclusive evidence for an effect which became inconclusive evidence against an effect after the data were transformed. As specified in the pre-registration document, due to differences in interpretation, we report the analyses using the transformed data here. Analyses using the raw data can be found in the supplementary materials available online.

We undertook a Bayesian mixed-ANOVA with VR condition (flat-screen; IA-VR) and violent content (non-violent; violent) as between-subjects factors and time (baseline, post-game) as a within-subjects factor. There was inconclusive evidence that VR condition did not moderate changes in hostile affect over time, $BF_{10} = 0.80$, and inconclusive evidence that violent content did not moderate changes in hostile affect over time $BF_{10} = 0.89$. There was substantial evidence against a three-way interaction, $BF_{10} = 0.19$. These results were similar when data were removed from participants who guessed the hypotheses of the study, with (IA-VR x time, $BF_{10} = 0.85$, violence x time, $BF_{10} = 0.93$, violence x IA-VR x time, $BF_{10} = 0.27$). These results were also similar when data were removed from any participants who expressed suspicion of the hypotheses of the study or the purpose of the Wasabi task (IA-VR x time, $BF_{10} = 0.24$, violence x time, $BF_{10} = 0.40$, violence x IA-VR x time, $BF_{10} = 0.29$).

Exploratory Analyses.

Relationship between aggressive cognition, hostile affect and aggressive behavior. One unanticipated finding of the present study was the difference between the effects of violent content upon aggressive cognition, hostile affect, and aggressive behavior. The GAM views aggression as a primarily cognitive phenomena with no need for an affective component (Ferguson & Dyck, 2012). For instance, Anderson and Dill (2000, p. 788) state that "…the danger in exposure to violent video games seems to be in the ideas they teach and not primarily in the emotions they incite in the player." As such, we were particularly interested in the relationship between aggressive behavior and aggressive cognition. Thus, although we did not pre-register this analysis, we believed it was important to explore the relationship between these measures given the divergent effects of violent content upon these measures in the present study.

Thus, we calculated pairwise Bayesian correlations between the measures. Initial analyses employed a non-informative prior with a stretched beta prior width of 1. As can be seen in Table 1, there is virtually no relationship between the number or proportion of aggressive words reported on the word stem completion task (indexing cognitive aggression) and the amount of wasabi administered by participants (indexing behavioral aggression). The word stem completion task was also unrelated to hostile affect. Finally, the wasabi task was unrelated to hostile affect. Altering the prior to an informative prior where the alternative hypothesis was specified to require a positive correlation between variables as would be predicted by the GAM theory resulted in similar evidence strengths (Wasabi/Total aggressive words $BF_{+0} = 0.06$; Wasabi/Proportion aggressive words, $BF_{+0} = 0.09$; Hostile affect/Total aggressive words, $BF_{+0} = 0.12$; Hostile affect/Total aggressive words, $BF_{+0} = 0.17$; Hostile affect/Wasabi $BF_{+0} = 0.21$). Evidence also remained in favour of the null for all effects when using an informative prior specifying positive associations as the alternative hypothesis with a beta prior width of 0.5 (Wasabi/Total aggressive words $BF_{+0} = 0.09$; Wasabi/Proportion aggressive words, $BF_{+0} = 0.13$; Hostile affect/Total aggressive words, $BF_{+0} = 0.19$; Hostile affect/Total aggressive words, $BF_{+0} = 0.25$; Hostile affect/Wasabi $BF_{+0} = 0.31$). No analysis yielded evidence in favour of a positive association between any of the measures of aggressive cognition, affect, and behavior.

Table 1
Relationship between measures of aggressive cognition, aggressive behavior and hostile affect (Study 1). 95% Credible Intervals appear in brackets.

	Total Number of Aggressive Words	Proportion of Aggressive Words	Hostile Affect (T2)
Wasabi Administered	$r = -0.04$ [−0.18, 0.10], $BF_{10} = 0.10$	$r = -0.00$ [−0.14, 0.14], $BF_{10} = 0.09$	$r = 0.06$ [−0.08, 0.20], $BF_{10} = 0.13$
Hostile Affect (T2)	$r = 0.03$ [−0.11, 0.17], $BF_{10} = 0.10$	$r = 0.05$ [−0.09, 0.19], $BF_{10} = 0.11$	

General Positive Affect. The general positive affect scale of the PANAS-X showed no evidence for skewness or kurtosis at either baseline (Skewness = 0.077, Kurtosis = 0.173) or at post-test (Skewness = −0.225, Kurtosis = −0.575). We therefore undertook a Bayesian mixed-ANOVA with VR condition (flat-screen; IA-VR) and violent content (non-violent; violent) as between-subjects factors and time (baseline, post-game) as a within-subjects factor on the raw positive affect scores. There was substantial evidence that violent content did not moderate changes in positive affect over time, $BF_{10} = 0.16$. However, there was substantial evidence that changes in positive affect over time were moderated by VR condition, $BF_{10} = 84.13$. Participants who played a flat-screen game had slightly lower positive affect between baseline ($M = 30.65$, $SD = 5.86$) to post-game tests ($M = 29.42$, $SD = 8.54$), a borderline small decline $d = 0.20$. However, participants who played an IA-VR game showed a small increase in positive affect between baseline ($M = 31.26$, $SD = 6.61$) and post-game tests ($M = 33.26$, $SD = 8.14$), $d = 0.33$. There was inconclusive evidence that a three-way interaction was less likely than a simple two-way interaction between VR condition and time, $BF_{10} = 0.33$. Exploratory analyses showed the IA-VR main effect was primarily due to changes in three positive emotion subscales. Participants who played a flat-screen game displayed stable levels of feeling active from baseline ($M = 3.02$, $SD = 1.00$) to post-game ($M = 2.89$, $SD = 1.09$), $d = 0.12$, participants who played a VR game experienced a small increase in feeling active between baseline ($M = 3.10$, $SD = 0.97$) and post-game (M = 3.42, SD = 1.09) measures, $d = 0.31$, $BF_{10} = 11.39$. Participants who played a flat-screen game displayed a small reduction in feeling inspired from baseline (M = 2.76, SD = 1.11) to post-game ($M = 2.38$, $SD = 1.20$), $d = 0.31$. In contrast, participants who played a VR game experienced stable feelings of being inspired between baseline ($M = 2.68$, $SD = 1.20$) and post-game ($M = 2.78$, $SD = 1.30$) measures, $d = 0.07$, $BF_{10} = 4.19$. Participants who played a flat-screen game displayed a small reduction in feeling excited from baseline (M = 3.31, SD = 0.89) to post-game ($M = 2.92$, $SD = 1.25$), $d = 0.31$. In contrast, participants who played a VR game experienced a small increase in feelings of being excited between baseline ($M = 3.21$, $SD = 1.04$) and post-game ($M = 3.47$, $SD = 1.10$) measures, $d = 0.21$, $BF_{10} = 135.88$.

General Negative Affect. The general negative affect scale showed high levels of kurtosis for baseline (Skewness = 1.81, Kurtosis = 4.11) and high levels of both skewness and kurtosis for post-game measures (Skewness = 2.11, Kurtosis = 5.92). Undertaking a square root transformation on the data did not normalize the data for either baseline (Skewness = 1.42, Kurtosis = 2.35) or post-game measures (Skewness = 1.57, Kurtosis = 2.80). As such, in accordance with the pre-registration protocol, we undertook a logarithmic transformation upon the data which normalized both the baseline (Skewness = 1.08, Kurtosis = 1.09), and post-game (Skewness = 1.15, Kurtosis = 1.00) measures. Most conclusions were similar when the untransformed data were used, although for one interaction the untransformed data showed substantial evidence for an interaction which became strong evidence for an interaction after the data were transformed. As specified in the pre-registration document, due to differences in interpretation, we report the analyses using the transformed data here. Analyses using the raw data can be found in the supplementary materials available online.

We undertook a Bayesian mixed-ANOVA with VR condition (flat-screen; IA-VR) and violent content (non-violent; violent) as between-subjects factors and time (baseline, post-game) as a within-subjects factor on logged general negative affect scores. Analyses showed that there was substantial evidence that violent content moderated changes in negative affect over time, $BF_{10} = 35.07$. Participants who played a non-violent game remained steady in their negative affect from baseline ($M = 1.11$, $SD = 0.10$) to post-game ($M = 1.10$, $SD = 0.12$), $d = 0.08$. However, participants who played a violent game had a small increase in negative affect from baseline ($M = 1.12$, $SD = 0.10$) to post game ($M = 1.16$, $SD = 0.13$), $d = 0.32$. There was substantial evidence that VR condition did not moderate changes in negative affect from baseline to post-game measures, $BF_{10} = 0.20$. There was substantial evidence against a three-way interaction $BF_{10} = 0.30$. Exploratory analyses showed that the increase in negative affect was due to participants who played a non-violent game showed small declines in fear from baseline ($M = 1.36$, $SD = 0.61$) to post-game ($M = 1.15$, $SD = 0.43$), $d = 0.31$, while participants who played a violent game experienced a small increase in fear from baseline ($M = 1.30$, $SD = 0.56$) and post-game (M = 1.55, SD = 0.93) measures, $d = 0.25$, $BF_{10} = 83.49$. Participants who played a non-violent game showed stable feelings of being scared from baseline ($M = 1.20$, $SD = 0.53$) to post-game ($M = 1.19$, $SD = 0.54$), $d = 0.02$, participants who played a violent game experienced a small increase in being scared from baseline ($M = 1.22$, $SD = 0.51$) and post-game (M = 1.50, SD = 0.79) measures, $d = 0.34$, $BF_{10} = 4.89$. All other negative affect subscales showed inconclusive evidence or in favour of the null.

Difficulty and Immersion. An error in the uploading of the preregistration document resulted in these analyses not being properly preregistered. We intended to use the single item measures of difficulty, immersion and novelty to investigate any effects of virtual reality on aggression. As we found no effect of virtual reality on any measure of aggression, and due to the error in the preregistration document, we did not conduct these analyses.

3. Discussion

We investigated the effects of IA-VR on aggressive cognition, behavior and hostile affect. Results showed no evidence that IA-VR alone, or in combination with violent content, affected aggressive cognition, behavior, or hostile affect. Although evidence for an interaction between IA-VR and violent content for aggressive behavior as measured by the administration of Wasabi was inconclusive ($BF_{10} = 0.34$), importantly, the evidence was *against* the presence of the interaction and trending toward substantial evidence for the null. Moreover, limiting analyses to those who did not guess the purpose of the wasabi task, in general, *strengthened* the evidence for null effects. IA-VR immersion did not amplify any relationship between violent game content and aggression.

The only effect of IA-VR was a small increase in general positive affect following gameplay. Exploratory analyses suggested this may be due to a small increase in feeling active and excited for the IA-VR group, contrasted against a small decline in feeling inspired and excited for participants who played flat-screen games. The increase in feeling active is unsurprising as IA-VR requires physical exertion. Increased excitement may be due to IA-VR novelty.

Although our data show increased accessibility of aggressive cognitions following violent gameplay, no effect on hostile affect or aggressive behavior occurred. These findings may be interpreted as supporting the growing evidence that violent media has a trivial-to-small influence on aggressive behavior. However, as there are potentially important narrative and game differences between the violent and non-violent games (e.g., goals, narration, humour, etc.), this result should be interpreted cautiously because, unlikely as it may be, we cannot rule out that differences between conditions may have been offset by differences between the stimuli.

Exploratory correlations show no relationship between aggressive cognition, hostile affect and aggressive behavior. This challenges a core assumption of the GAM: that activating and rehearsing aggressive cognitions primes aggressive behaviors. While we utilised instruments which have been validated and widely employed, it is also possible that these measures may have some validity problems. Both the PANAS-X and word-stem completion tasks have been previously validated (Carnagey & Anderson, 2005; Watson & Clark, 1999). Similarly, the Wasabi task, a variant of the hot-sauce task has been used to validate other aggressive behavior measures such as the Taylor Aggression Paradigm (Chester & Lasko, 2019). Nonetheless, basic psychometric validation of the scales used to assess aggressive cognition and emotion may be beneficial to help establish the construct validity of these scales and the validity of the present findings.

We did observe a small increase in negative affect for participants who played the violent games. However, exploratory analyses revealed these effects to be due to increased feelings of being scared and afraid following gameplay. As both violent games were wave-based zombie survival shooters, the horror content of the games may have been responsible for these increases in participants' fear.

One limitation of the present study is that the violent games employed to assess differences between flat-screen and IA-VR content were not equivalent. Although the games are both rated M 17+ by the ESRB for Blood and Gore, Drug references, Intense Violence, and Strong Language, differences between the narrative context or gameplay may have contributed to a lack of differences in aggression. By the time Study 1 was completed, the games catalogue for IA-VR systems had expanded to include Doom-VFR. This enabled us to conduct Study 2 to compare outcomes for players playing Doom in IA-VR and non IA-VR conditions; addressing this limitation.

Study 2.

Study 1 found virtually no evidence to support the notion that IA-VR increased aggressive cognition, affect, or behavior. Here we sought to replicate these findings by directly testing whether a VR, compared to Non-VR, version of the same game (Doom and Doom-VFR) would cause changes in participants' aggressive affect, cognition, or behavior.

4. Method

Participants & Design.

In this study, we report all measures, manipulations, and exclusions. Ninety-six participants (49 male) recruited from a university campus in New Zealand participated in the study. We undertook a pre-registered protocol with clear stopping rules and had aimed to collect 100 usable participants for this study. Our pre-registration including stopping rules can be accessed at: https://aspredicted.org/ce8nc.pdf However, we terminated data collection due to data collection interruptions associated with the global COVID-19 pandemic. Specifically, we terminated data collection when reaching 96 usable participants due to lockdowns in New Zealand. Although these lockdowns lifted sometime later, the country remained at Alert Level 1 or higher since the cessation of the initial lockdown. As IA-VR involves close contact with participants' and the use of a communal head and facial display, we determined the safest course of action was to cease data collection at 96 participants. As we had collected the majority of the participants for this study, we felt comfortable that our data were an adequate test of our hypothesis and terminated data collection. No analyses were performed prior to the decision to permanently halt data collection.

As pre-registered, participants who scored five or more on the Generalised Anxiety Disorder 2 item scale (GAD-2; $n = 2$) were excluded from participating in the study due to the horror content of the games and were excluded as specified in the pre-registration document. A further 2 participants ceased gameplay due to the horror content after playing for less than 10 min and their data were discarded in accordance with our pre-registration. Additionally 3 participants were replaced due to failing to follow experimenter instructions. As these participants had received a participant number prior to starting the experiment, data from these participants was replaced with data from a new participant with an R after their participant number to denote them as replacements. Due to the termination of data collection resulting from the

COVID-19 pandemic, some replacement participants were unable to be collected, resulting in a final sample of $n = 96$ participants.

Participants were randomly allocated to a 2 (virtual reality: flat screen; HTC Vive) between-subjects experimental design.

Sensitivity Power Analysis. Sensitivity power analysis assuming an alpha level of 0.05 and a power of 80% revealed that our sample was sufficient to reliably detect moderately sized differences between the means of two of our groups at an effect size of $d = 0.51$.

Materials and Measures.

All measures were the same as Study 1, and the procedure was identical. Only 3 changes were implemented for Study 2. First, to control for potential confounds in the experimental materials, we compare a violent flat-screen (DOOM) and a violent IA-VR-game (DOOM VFR) which were much more similar than in Study 1. Although small differences still exist between these games to accommodate VR play, the games employ the same monsters, weapons, objectives and environments, and both games have the same ESRB rating. This eliminates much of the variability between the two violent games. Moreover, this gives our study high ecological validity by comparing a market-available VR and non-VR version of the same game. Second, as our primary interest was whether violent IA-VR games increase aggression relative to violent flat-screen games, we do not include a non-violent control group. Third, we included a number of single item measures to determine participants' self-reported play experiences – specifically how difficult, immersive, novel, frustrating, engaging, boring, fast paced, humorous, scary, violent, relaxing, exciting, and realistic the game was (1, not at all – 7, extremely) and how much they would want to play the game again.

5. Results

Confirmatory analyses.

The complete dataset is publicly available (https://osf.io/t9yrc/). Data were analysed in accordance with the pre-registration protocol document. To reduce potential bias, initial analyses were conducted by an analyst blind to the coding of the conditions. All analyses were conducted using Bayesian methods. For all Bayesian *t*-tests we employed uninformative priors with default Cauchy widths of 0.707.

Aggressive Cognition. As per Study 1, we calculated both the total number of aggressive words reported and the total number of aggressive words reported expressed as a proportion of the total number of word stems attempted. These approaches yielded identical conclusions. Data for the word stem completion task showed normal values for skewness and kurtosis for both total number of aggressive words reported (skewness = 0.99, kurtosis = 0.01) and proportion of aggressive words reported (skewness = 0.62, kurtosis = 0.76). We ran Bayesian t-tests on the number of aggressive words reported and the proportion of aggressive words reported with VR condition (flat-screen; IA-VR) as the independent factor. In accordance with the findings of Study 1, there was substantial evidence that participants' reported similar numbers of aggressive words in total when they played the violent flat-screen ($M = 6.04$, $SD = 3.04$) and violent IA-VR games ($M = 5.92$, $SD = 2.58$), $BF_{10} = 0.22$, $d = 0.04$. Participants also reported similar proportions of aggressive words in the flat-screen ($M = 17.0\%$, $SD = 7.1\%$) and IA-VR conditions ($M = 17.5\%$, $SD = 7.1\%$), $BF_{10} = 0.23$, $d = 0.07$. Eleven participants guessed the hypothesis of the study, and 65 reported suspicion of the hypothesis of the study. Twelve participants guessed the purpose of the Wasabi task while a further 42 participants were suspicious of the Wasabi task. Limiting analyses to participants who did not guess the hypothesis of the study or the purpose of the Wasabi task did not substantially alter the results (Total aggressive words, $BF_{10} = 0.31$, Proportion of aggressive words, $BF_{10} = 0.24$). Only 11 participants were coded as not reporting some suspicion of either the hypothesis or the Wasabi task, leaving too few to analyse reliably. In no analyses did we observe evidence supporting the effect of IA-VR on aggressive cognition.

Aggressive Behavior. Data for the wasabi task showed high levels of both skewness (2.95) and Kurtosis (10.61). A square root transformation did not normalize the data (Skewness = 1.09, Kurtosis = 2.28). As such, in accordance with the pre-registration documentation, we undertook a logarithmic transformation on the data which normalized the data (Skewness = −0.19, Kurtosis = 0.43). The conclusions favoured the null for both the raw and logarithmically transformed data, though the strength of evidence was changed by the logarithmic transformation. In accordance with the pre-registration document, we report the logarithmically transformed data here with results for the raw data reported in footnotes.

In agreement with the findings from Study 1, a Bayesian *t*-test with VR condition (flat-screen; IA-VR) as the independent factor revealed substantial evidence in favour of the null hypothesis. Bayes factors showed substantial evidence for null effect of the IA-VR on the natural logarithm of the amount of Wasabi administered by participants in the flat-screen ($M = -0.39$, $SD = 0.81$) and IA-VR conditions ($M = -0.23$, $SD = 1.12$), $BF_{10} = 0.28$[1] $d = 0.16$. Limiting the analyses to participants who did not guess the Wasabi task was a measure of aggression weakened the evidence for the null slightly, $BF_{10} = 0.50$, and eliminating all participants who expressed some suspicion about the wasabi task being a measure of any form of negative behavior split the difference between these two analyses in terms of strength of evidence, $BF_{10} = 0.32$. Eliminating participants who guessed the hypothesis of the study and participants who guessed the purpose of the Wasabi task yielded similar results, BF10 = 0.40. Only 11 participants were coded as not reporting some suspicion of either the hypothesis or the Wasabi task, leaving too few to analyse reliably. In no analyses did we observe evidence supporting the effect of IA-VR on aggressive behavior.

Hostile affect. Initial normality checks indicated that baseline data on the PANAS-X hostility subscale was normal (Skewness =0.77, Kurtosis =1.49). However, post-game data on the hostility subscale showed normal Skewness (1.53) but slightly heavy-tailed Kurtosis (2.75). A Square root transformation normalized these data (Skewness =1.14, Kurtosis =1.52). However, analyses using the transformed and raw data produced qualitatively similar results leading to identical interpretations. Thus, in accordance with the pre-registration document, we report analyses on the raw data here.

We undertook a Bayesian mixed-ANOVA with VR condition (flat-screen; IA-VR) as a between-subjects factors and time (baseline, post-game) as a within-subjects factor. Replicating the results of Study 1, there was substantial evidence that VR condition did not moderate changes in hostile affect over time, $BF_{10} = 0.27$. Eliminating participants who guessed the hypothesis of the study and participants who guessed the purpose of the Wasabi task yielded weaker evidence, though the evidence was still in favour of the null ($BF_{10} = 0.57$). In no analysis did we observe evidence in favour of IA-VR condition increasing hostile affect.

General Positive and Negative Affect. Bayesian repeated measures ANOVA's using time as a within-subjects factor and VR condition as a between-subjects factor revealed inconclusive evidence for the null effect of the interaction between IA-VR condition and time on positive affect, $BF_{10} = 0.48$ and on negative affect $BF_{10} = 0.48$.

Relationship between aggressive cognition, hostile affect and aggressive behavior. One unanticipated finding of Study 1 was the weak associations between aggressive cognition, hostile affect, and aggressive behavior. Here, we undertook pre-registered pairwise Bayesian correlations between the measures to replicate the findings of Study 1 employing a default non-informative prior with a stretched beta prior width of 1. Our findings replicated the results of Study 1. As can be seen in Table 2, there is virtually no relationship between the number or proportion of aggressive words reported on the word stem completion task (indexing cognitive aggression) and the amount of wasabi

[1] Using raw instead of logarithmically transformed data weakened the strength of the evidence, though evidence remained in favour of the null, $BF_{10} = 0.65$.

Table 2
Relationship between measures of aggressive cognition, aggressive behavior and hostile affect (Study 2). 95% Credible Intervals reported in Brackets.

	Total Number of Aggressive Words	Proportion of Aggressive Words	Hostile Affect (T2)
Wasabi Administered	$r = -0.05$ [−0.25, 0.15], $BF_{10} = 0.14$	$r = 0.03$ [−0.17, 0.22], $BF_{10} = 0.13$	$r = -0.155$ [−0.340, 0.047], $BF_{10} = 0.39$
Hostile Affect (T2)	$r = 0.09$ [−0.12, 0.28], $BF_{10} = 0.18$	$r = 0.06$ [−0.14, 0.25], $BF_{10} = 0.15$	

administered by participants (indexing behavioral aggression). The word stem completion task was also unrelated to hostile affect. Finally, the wasabi task was unrelated to hostile affect. Altering the prior to an informative prior where the alternative hypothesis was specified to require a positive correlation between variables as would be predicted by the GAM theory resulted in similar evidence strengths (Wasabi/Total aggressive words $BF_{+0} = 0.09$; Wasabi/Proportion aggressive words, $BF_{+0} = 0.16$; Hostile affect/Total aggressive words, $BF_{+0} = 0.28$; Hostile affect/Total aggressive words, $BF_{+0} = 0.22$; Hostile affect/Wasabi $BF_{+0} = 0.05$). Evidence also remained in favour of the null for all effects when using an informative prior specifying positive associations as the alternative hypothesis with a beta prior width of 0.5 (Wasabi/Total aggressive words $BF_{+0} = 0.13$; Wasabi/Proportion aggressive words, $BF_{+0} = 0.23$; Hostile affect/Total aggressive words, $BF_{+0} = 0.42$; Hostile affect/Total aggressive words, $BF_{+0} = 0.32$; Hostile affect/Wasabi $BF_{+0} = 0.08$). No analysis yielded evidence in favour of a positive association between any of the measures of aggressive cognition, affect, and behavior.

These results call into question some basic assumptions of the General Aggression Model – in particular whether the constructs of cognitive and affective aggression are meaningfully related. As noted in Study 1, we employed instruments which have been validated and widely employed. However, further validation of these instruments would increase our confidence in the findings of the present manuscript.

Exploratory analyses.

Exploratory analyses of the participants' experiences with the game yielded 6 self-reported variables which showed evidence for a difference between IA-VR and Flat-Screen games. Table 3 shows these effects.

One potential explanation for the null findings in Study 2 was that these self-reported differences might be suppressing an effect. When we included these variables in a Bayesian ANCOVA, there was still substantial evidence for the null effect of VR condition on the total number of aggressive words reported ($BF_{10} = 0.26$), inconclusive evidence (approaching substantial evidence) for the null for the proportion of aggressive words reported $BF_{10} = 0.32$, and substantial evidence for a null effect of VR condition on aggressive behavior $BF_{10} = 0.27$. There was also anecdotal evidence (trending toward substantial evidence) for a lack of interaction between time and VR condition on hostile mood, $BF_{10} = 0.39$. It does not appear that self-reported differences in the experiences of the two games was responsible for the lack of observed differences between IA-VR and flat-screen conditions.

General Discussion.

In two pre-registered studies we investigated the effects of violent IA-VR games on aggressive cognitions, affect and behavior. These findings are the first examination of IA-VR video games, and the extent they might moderate relationships between game content and aggression. To the extent that the instruments we employ are valid measures of aggression, our results suggest that the violent IA-VR games we tested did not increase aggressive cognitions, emotions or behaviors. In contrast, in Study 1, IA-VR games appeared to make users happier, via a combination of increased feelings of activity and excitement. In Study 2, using a VR and non-VR version of the same game we found no effect of IA-VR on positive or negative affect. Taken together the results suggest that violent IA-VR games may not increase users' aggressiveness compared to violent flat-screen games, though further replications with a wider range of games, methods, measures, and materials are required to confirm these findings.

Table 3
Variables showing positive evidence for a difference in self-reported experience between IA-VR and Flat-Screen games in Study 2.

Variable	Flat-Screen	IA-VR	Cohens *d*, Bayes Factor
Immersive	4.51 (1.20)	5.33 (1.23)	$d = 0.67$, $BF_{10} = 22.25$
Novel	4.04 (1.27)	5.33 (1.35)	$d = 0.98$, $BF_{10} = 2851.81$
Violent	4.79 (1.56)	3.96 (1.61)	$d = 0.52$, $BF_{10} = 3.71$
Exciting	4.32 (1.45)	5.41 (1.61)	$d = 0.71$, $BF_{10} = 38.19$
Realistic	2.06 (1.23)	3.59 (1.83)	$d = 0.98$, $BF_{10} = 2190.89$
Would Play Again	3.85 (1.79)	5.51 (1.72)	$d = 0.95$, $BF_{10} = 1470.18$

One unanticipated and important finding from both studies was that there was virtually no relationship between the employed measures of aggressive cognition, hostile affect, and aggressive behavior (as measured by the Wasabi task). The fact that in both studies we observed very weak correlations between these measures (and Bayes Factors consistently in favour of a null-effect) calls into question some basic assumptions of the GAM – that is, that aggressive affect, cognition, and behavior are interrelated. Basic examinations of the GAM may be required. Further psychometric validation of the commonly employed measures of aggression may also be useful to ensure that they are successfully measuring the constructs of interest.

Our studies have several limitations. First, we employ the Wasabi task, a slightly modified version of the Hot Sauce task (Fischer et al., 2010). The hot sauce task has been used to validate other behavioral aggression measures such as the Taylor Aggression Paradigm (Chester & Lasko, 2019). Unlike the Taylor Aggression Paradigm, the Wasabi and Hot-Sauce methods do not suffer from multiple scoring methods which could encourage unidentified researcher degrees of freedom (e.g., Elson, Mohseni, Breuer, Scharkow, & Quandt, 2014). Nonetheless, the modification of the task from hot-sauce to Wasabi may affect the validity of this measure, and the results should thus be replicated with other measures of aggression to determine whether the lack of effect holds across measures as well as stimulus materials.

A second limitation of the present study is that a high proportion of participants either guessed the hypothesis of the study or the purpose of the Wasabi task. While this is not unusual when employing such tasks, it does create a potential social desirability bias and further calls into question the use of such measures in media violence research. Our analyses were similar when suspicious participants were included or excluded, but it is possible that there are individual differences between people who are more likely to hypothesis guess which might impact the results. Development of novel methods to assess aggression which are less obtrusive would be a worthy objective. Validation of standardised funnelled debriefing for these studies to detect hypothesis guessing and deception detection would also be a valuable goal.

A third potential limitation of the present study is the relatively small sample sizes employed. Virtual reality research is time and resource intensive, making larger sample sizes impractical, especially for smaller laboratories. We have partially accounted for the small sample sizes by employing Bayesian analyses which are more robust to small samples (Van De Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & Van Loey, 2015). We observed virtually no evidence in either of our studies in favour of an effect of IA-VR on any measure of aggressiveness (one analysis excluding all suspicious participants suggested an effect of VR on the number of aggressive words, but not on the proportion of aggressive words reported, suggesting contradictory conclusions). When taken together, the predominance of our results suggest substantial accumulated evidence for a null effect. Nonetheless, we believe further work employing larger samples is a worthy goal to determine the replicability of the present findings.

In sum, our study provides some early evidence that violent IA-VR games may not be associated with increases in aggression compared to flat-screen games. Further replication of these findings with different games and aggression measures will be important to validate these findings. Our study also shows surprisingly small associations between the measures of aggressive cognition, aggressive behavior, and hostile affect. This suggests that basic validation of the GAM may be required.

Open Practices.

The pre-registration for Study 1 is available at: https://aspredicted.org/d9a3y.pdf. The pre-registration for Study 2 is available at: https://aspredicted.org/ce8nc.pdf. Data and questionnaires for both studies are openly available at https://osf.io/t9yrc/

Funding

Supported by the Marsden Fund Council from Government funding, managed by Royal Society Te Apārangi; MAU1804.

Appendix A. Supplementary analyses

Supplementary analyses.

References

Adachi, P. J., & 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.

Anderson, C. A., Bushman, B. J., Bartholow, B. D., Cantor, J., Christakis, D., Coyne, S. M., . . . Green, C. S. (2017). Screen violence and youth behavior. *Pediatrics, 140* (supplement 2), S142-S147.

Anderson, C. A., & Carnagey, N. L. (2004). Violent evil and the general aggression model. *The social psychology of good and evil*, 168–192.

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.

Arriaga, P., Esteves, F., Carneiro, P., & Monteiro, M. B. (2008). Are the effects of unreal violent video games pronounced when playing with a virtual reality system? *Aggressive Behavior: Official Journal of the International Society for Research on Aggression, 34*(5), 521–538.

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.

Chester, D. S., & Lasko, E. N. (2019). Validating a standardized approach to the Taylor aggression paradigm. *Social Psychological and Personality Science, 10*(5), 620–631.

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2011). The general aggression model: Theoretical extensions to violence. *Psychology of Violence, 1*(3), 245.

Donker, T., van Straten, A., Marks, I., & Cuijpers, P. (2011). Quick and easy self-rating of generalized anxiety disorder: Validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. *Psychiatry Research, 188*(1), 58–64.

Drummond, A., Sauer, J. D., & Ferguson, C. J. (2020). Do longitudinal studies support long-term relationships between aggressive game play and youth aggressive behaviour? A meta-analytic examination. *Royal Society Open Science, 7*(7), 200373.

Drummond, A., Sauer, J. D., & Garea, S. S. (2018). The infamous relationship between violent video game use and aggression – Uncharted moderators and small effects make it a far cry from certain. In C. J. Ferguson (Ed.), *Video game influences on aggression, cognition, and attention*. Washington DC: Spring.

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.

Entertainment Software Association (2017). Essential facts about the computer and video game industy. Retrieved from http://essentialfacts.theesa.com/.

Farrar, K., & Krcmar, M. (2006). Measuring state and trait aggression: A short, cautionary tale. *Media Psychology, 8*(2), 127–138.

Ferguson, C. J., & Dyck, D. (2012). Paradigm change in aggression research: The time has come to retire the general aggression model. *Aggression and Violent Behavior, 17*(3), 220–228.

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*, 174–178.

Fischer, P., Kastenmüller, A., & Greitemeyer, T. (2010). Media violence and the self: The impact of personalized gaming characters in aggressive video games on aggressive behavior. *Journal of Experimental Social Psychology, 46*(1), 192–195.

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., … Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry, 5*(8), 625–632.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168.

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).

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Krcmar, M., & Farrar, K. (2009). Retaliatory aggression and the effects of point of view and blood in violent video games. *Mass Communication and Society, 12*(1), 115–138.

Lin, J. H. T. (2017). Fear in virtual reality (VR): Fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. *Computers in Human Behavior, 72*, 350–361.

Lull, R. B., & Bushman, B. J. (2016). Immersed in violence: Presence mediates the effect of 3D violent video gameplay on angry feelings. *Psychology of Popular Media Culture, 5*(2), 133.

Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., & Tunbridge, E. M. (2019). It feels real: Physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology, 33*(10), 1264–1273.

Mathur, M. B., & VanderWeele, T. J. (2019). Finding common ground in meta-analysis "wars" on violent video games. *Perspectives on Psychological Science, 14*(4), 705–708.

Peterson, S. M., Furuichi, E., & Ferris, D. P. (2018). Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading. *PLoS One, 13*(7), Article e0200306.

Prescott, A. T., Sargent, J. D., & Hull, J. G. (2018). Metaanalysis of the relationship between violent video game play and physical aggression over time. *Proceedings of the National Academy of Sciences, 115*(40), 9882–9888.

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.

Sauer, J. D., & Drummond, A. (2020). Boundary conditions for the practical importance of small effects in long runs: A comment on Funder and Ozer (2019). *Advances in Methods and Practices in Psychological Science, 2515245920957607*.

Sauer, J. D., Drummond, A., & Nova, N. (2015). Violent video games: The effects of narrative context and reward structure on in-game and postgame aggression. *Journal of Experimental Psychology: Applied, 21*(3), 205–214.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.

Tamborini, R., Eastin, M. S., Skalski, P., & Lachlan, K. (2004). Violent virtual video games and hostile thoughts. *J. Broad. & Elec. Media, 48*, 335.

Van De Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & Van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology, 6*(1), 25216.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*.