Hero Violence

Joe Hilgard

June 7, 2017

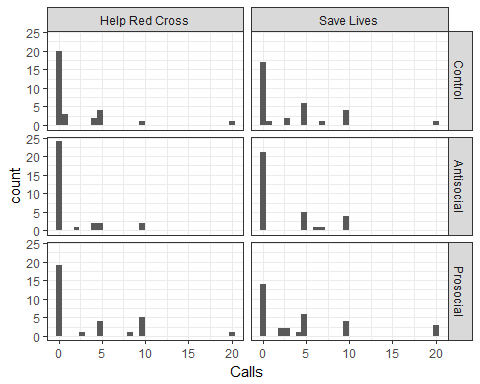
## Sample

An initial data collection included 216 subjects, 191 of which were successfully intercepted by the confederate as they left the laboratory. During data collection, the Spring semester ended before data collection was completed. We looked at the data to examine whether it was trending in the hypothesized direction, with the idea that we would continue data collection as planned if it was, and would drop data collection if it was not. The data was indeed trending in the hypothesized direction, so we picked up data collection as originally planned in the following Fall semester. This natural pause in data collection and evaluation of whether or not to continue collecting data may or may not influence Type I error rates. Additionally, we decided to exclude a semester's worth of participants from one study site (*n* = 31, 21 of which were intercepted), as the confederate used for only that semester did not correctly collect data as was instructed. To support this exclusion of the data, we found the participants solicited by this confederate were statistically less likely to volunteer than other participants, chi-square(1) = 4.02, *p* = .045. This left us with a final sample of 204 subjects, 189 of which were successfully intercepted by the confederate. This left us with a final sample of 204 subjects, 189 of which were successfully intercepted.

Our primary analyses concern this final sample. However, to explore the robustness of our results, we also analyze the data including those 31 excluded participants.

## Distribution

Initial inspection of the data revealed pronounced non-normality (Figure 1). Many subjects did not offer to make any calls, and among thosewho did, the number of calls offered was strongly right-skewed and favored multiples of 5.



This non-normality creates ambiguity in what might be the most appropriate model to fit. Although ANOVA is commonly used and fairly robust to outliers, its assumptions would seem to be violated by the pronounced non-normality of the residuals (see e.g. Glass, Peckham, and Sanders, 1972). Therefore, we fit alternative models to explore the robustness of the result and to characterize the analytic ambiguity of the results. In total, we fit: 1) ANOVA 2) logistic regression, testing whether condition affected the odds of volunteering to make any calls 3) a chi-squared test, again testing whether condition affected the odds of volunteering to make any calls 4) a zero-inflated negative binomial, which accounts for the frequent zero-responses and the strong right skew, and 5) a non-parametric Kruskal-Wallis test for differences in the median. Results for each model are reported separately, then synthesized and summarized at the end.

### ANOVA

ANOVA indicated a significant effect of game, *F*(2, 183) = 3.2, *p* = .043. Neither the effect of request nor the Game × Request interaction was statistically significant, *F*(1, 183) = 1.27, *p* = .262 and *F*(2, 183) = 0, *p* = .996, respectively.

Collapsing across levels of request, a contrast between the prosocial-violence and antisocial-violence conditions was statistically significant, *t*(122) = 2.48, *p* = .015, *b* = 2.02 [.40, 3.63]. Neither game significantly differed from control.

### Logistic GLM

To test whether the game influenced the odds of volunteering, we collapsed observations to a binomial outcome (0 = did not volunteer, 1 = volunteered). A logistic GLM was fit to test for effects.

The effect of game was not statistically significant, but not especially far from significance either, *chisq*(2) = 5.09, *p* = .079. Effects of request and the Game x Request interaction were not statistically significant, *chisq*(1) = 0.85, *p* = .358 and *chisq*(2) = 0.11, *p* = .948, respectively.

Again, the contrast between the prosocial-violent and antisocial-violent games was statistically significant, *z* = 2, *p* = .029, *OR* = 2.31 [1.1, 4.96]. However, neither game significantly differed from control.

## Chi-squared test

A chi-squared test was used to test whether the probability of volunteering at all differed across game conditions (collapsing across levels of request). The omnibus test for game type was not significant, but not far from significance, chi-square(2) = 4.97, *p* = .083. As in other analyses, the contrast between prosocial-violent and antisocial-violent was just significant, chi-square(1) = 4.07, *p* = .044, although neither game significantly differed from control.

## Zero-inflated negative binomial

A zero-inflated negative binomial was then fit to the data. This model has two parts: The zero-inflation model, which estimates whether participants volunteer at all, and the count model, which estimates how many calls they offer when they do volunteer.

Because we did not have reason to expect the game would influence only the zero-inflation model or only the count model, we tested the overall effect of game on both parameters simultaneously with likelihood-ratio chi-square tests. Within the zero-inflated negative binomial model, the effect of game was ambiguous, neither statistically significant nor particularly far from significance (*chisq*(4) = 7.92, *p* = .095). Effects of request (*chisq*(2) = 4.27, *p* = .118) or the Game × Request interaction (*chisq*(4) = 0.96, *p* = .916) were not significant. The overdispersion parameter was highly significant (*z* = 4.61, *p* < .001), supporting the use of the negative binomial over a zero-inflated Poisson distribution.

A pairwise contrast between the prosocial-violent and antisocial-violent games yielded a significant effect of game, *chisq*(2) = 6.67, *p* = .036. Tests of individual parameters found a significant difference on the zero-inflation parameter (*z* = -2.17, *p* = .030) but not on the count parameter (*z* = 1.30, *p* = .193). Again, neither game differed from control.

## Kruskal-Wallis test

The Kruskal-Wallis test is a nonparametric test of group medians. Because the test is only applicable to one-way ANOVA designs and the primary hypothesis regarded effects of game, we collapsed across levels of request and tested for an effect of game.

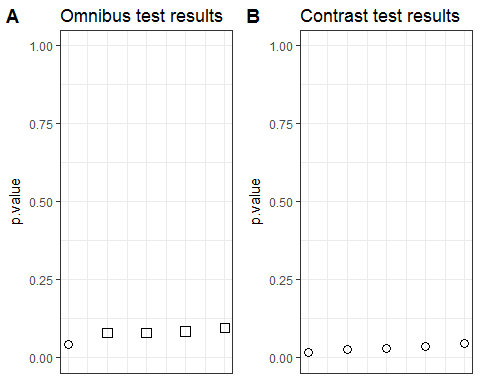
The omnibus test of different group means was only marginally significant, *chisq*(2) = 5.09, p = .079.

The pairwise contrast between the antisocial and prosocial violent games was statistically significant, *chisq*(1) = 4.97, *p* = .026. However, neither game significantly differed from control, *p* = .224, .262.

# Describing uncertainty

## Uncertainty across models

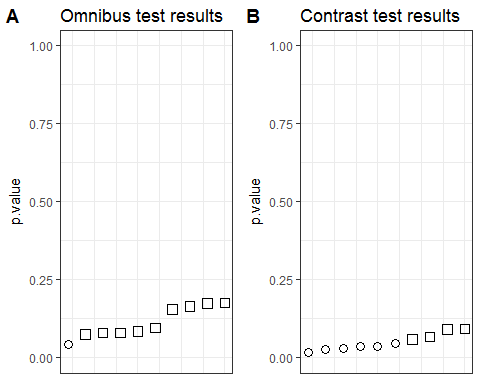
Figure 2A shows the distribution of omnibus test *p*-values across the five models. Figure 2B shows the distribution of pairwise prosocial/antisocial game contrast *p*-values across the five models.



Across models, the omnibus test is generally not statistically significant (median *p* = 0.079), reaching a minimum p-value of 0.043 in the ANOVA model. The pairwise contrast, on the other hand, is generally statistically significant. Its *p*-values range from .015 (ANOVA) to .044 (chi-squared test), with median *p* = 0.029.

## Uncertainty across datasets

In the analyses presented above, 31 participants were excluded over concerns regarding quality of the research assistants' application of the methods. [GIBSON, LUEKE, AM I SAYING THIS RIGHT?] One may wonder how robust the results are to the inclusion of these subjects. Figure 3 shows the variability in the *p*-value across this decision to include or exclude. Six parameters in two datasets yields twelve *p*-values. Figure 3A shows the distribution of omnibus test *p*-values, and Figure 3B shows the distribution of pairwise contrast *p*-values across the five models.



Together, these figures illustrate the uncertainty in the *p*-value across datasets. Again, we note that only one analysis yielded a significant omnibus test. Additionally, some of the pairwise contrasts are no longer statistically significant when including the 31 subjects of questionable quality. Naturally, bad data should be excluded. Still, to the extent that this decision was made *ad hoc*, it is useful to note their influence on the *p*-value.

# Bayesian ANOVA

One last way to consider the results would be through a Bayesian ANOVA. Like the frequentist ANOVA presented above, this model suffers from violated assumptions of normally distributed residuals. However, it does have the advantage of appraising the evidence in terms of continuous odds rather than the dichotomy of statistical significance. Effects were tested using the anovaBF() function in the BayesFactor package (Morey & Rouder, 2015) using a Cauchy prior with scale parameter 0.4, consistent with effects typical of social psychology.

These ANOVAs indicated ambiguous evidence regarding an omnibus effect of game, 1.15:1 odds in favor of H1 over H0. Odds in favor of the prosocial/antisocial game contrast were stronger, but still marginal at 2.96:1 in favor of H1 over H0. When including the 31 subjects of questionable quality, these ANOVAs indicated 1.94:1 odds against an omnibus effect of game and 1.44:1 odds in favor of a prosocial/antisocial game contrast.

This Bayesian approach indicates that the results are ambiguous and that neither H1 nor H0 is a clear winner, consistent with the other analyses presented above.

# Summary

We generally did not detect a significant overall effect of game on prosocial behavior. Still, pairwise contrasts suggested a significant difference in behavior following the prosocial-violent as compared to the antisocial-violent game. However, this contrast should be interpreted with caution given that neither violent game significantly differed from the control game.

Considerable uncertainty remains. The type I error rate will be influenced by our decision to collect additional data after the initial collection. The decision to exclude subjects is justifiable, but not based on any *a priori* rule, creating another researcher degree of freedom. The prosocial-violence vs. antisocial-violence contrast is one of three potential contrasts that consider the effects of game. Finally, the results are a little sensitive to the particular model applied. Whereas a typical manuscript might conceal the ambiguity by presenting just one analysis, here we find it most helpful to show all possible analyses.

This uncertainty should not be mistaken for an argument that games do not influence prosocial behavior. Although statistical significance is sporadic and sensitive to choice of model, contrast, or inclusion rules, it is generally not the case that the results constitute clear support for the null hypothesis.

The results of this study are nonetheless helpful in estimating the effects of antisocial-violent and prosocial-violent games. Although it is unclear whether these games have effects that differ from that of a control game, we did learn that if such effects exist, they are not large and obvious. Furthermore, we are able to rule out some effects of opposite direction. The data are inconsistent with antisocial-violent games *promoting* prosocial behavior or prosocial-violent games *inhibiting* prosocial behavior.

In conclusion, despite collection of a moderately-sized sample, the results are ambiguous regarding whether prosocial-violent and antisocial-violent games affect prosocial behavior as compared to a nonviolent game. Some support for a difference between antisocial and prosocial violent games was found, but this difference was sensitive to the choice of model and dataset. Future research is encouraged to test the effects of prosocial, antisocial, violent, and nonviolent games on prosocial behavior. This research program would be facilitated by the development of more sensitive and model-friendly measures of prosocial behavior.