The Killer: Moral Choice in Virtual Environments

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Justin H. Chang, B.A.

Graduate Program in Communication

The Ohio State University

2018

Thesis Committee

Brad J. Bushman, Advisor

Richard Huskey

Abstract

Most popular entertainment media contains content that consumers would consider immoral if it occurred in real life. This study compares two models that explain this phenomenon, the Moral Disengagement Model and the Model of Intuitive Morality and Exemplars, and tests them in the context of a video game containing a moral dilemma. 208 participants completed an experiment in which they were given the option to kill or spare a virtual agent in a video game. The moral justification for killing was manipulated, and the participants' moral foundations were measured to determine which model better predicted participant behavior. The results showed that neither model significantly predicted behavior; however, the data do show that people will anthropomorphize virtual agents with very little visual realism in video games, that killing these unrealistic agents still leads to feelings of guilt, and that seemingly simple moral dilemmas might activate several moral foundations.

Vita

2008 ……Juanita High School

2014 ……B.A. Political Science, Brigham Young University

2016 to present ……...G.A., The Ohio State University

Fields of Study

Major Field: Communication

Table of Contents

List of Tables

List of Figures

Introduction

Many of our most cherished stories revolve around immoral acts. Hamlet is a story of murder, incest, and deceit; Star Wars shows the genocide of entire planets; and Breaking Bad follows a school teacher's descent into the criminal world of the drug trade. Yet these and other stories are highly-praised and widely enjoyed, even though their narratives use characters, including protagonists, violate moral norms. Understanding why we take pleasure in stories that contain immoral acts in an ongoing investigation. This study seeks to contribute to this investigation by comparing two competing theories that explain the phenomenon: (1) the Model of Intuitive Morality and Exemplars (Tamborini, 2011) and (2) the Moral Disengagement Model (Hartmann & Vorderer, 2010). Although both theories explain the same phenomenon, they propose different processes and make different predictions. To my knowledge, the two theories have never been directly tested against each other. This study will test the predictive power of both theories by applying them to the same, tightly-controlled video game stimulus.

Background: Perspectives on Morality

The two models this study will examine come from competing theories of morality. Broadly speaking, there are three prominent theories of morality—a rationalist perspective (Kohlberg, Levine, & Hewer, 1983), an intuitive perspective (Zajonc, 1980), and an integrated dual-process model that seeks to synthesize the previous two perspectives (Greene & Haidt, 2002). I will provide a brief overview of the development and assumptions of these perspectives below.

Rationalist Perceptive

The rationalist perspective is based in Kohlberg's work on moral development (1976) that proposes a 6-stage hierarchy of moral reasoning. According to this theory, people progress from lower to higher stages as they develop cognitively and this 6-stage hierarchy describes a fundamental pattern of moral reasoning that should be observed across cultures (Kohlberg, 1971). However, Kohlberg (1971) clarified that this theory described moral *thought* and not necessarily moral *action*, noting that it is possible to act immorally despite understanding morality at a high level.

Later work in Social Cognitive Theory (Bandura, 1991) noted that Kohlberg's hierarchy poorly predicted moral judgments and had little empirical support. Still, the separation of moral thought and moral action was used as a framework to advance the concept of moral agency as a self-regulatory process. According to Bandura (1991), people refrain from committing immoral acts because they monitor their own behavior and circumstances, make moral judgments about potential behaviors, and self-censure when a potential behavior would (1) yield negative consequences such as social condemnation, (2) conflict with their moral standards and lead to

self-condemnation, or both. Put another way, people make conscious moral evaluations of their circumstances, predict the outcomes of potential behaviors, and act according to those evaluations and predictions.

But as both Kohlberg and Bandura pointed out, people still act in ways that violate their moral standards. Social Cognitive Theory seeks to explain this inconsistency by arguing that the self-regulatory process must be "activated" in order to function and that self-sanctions can be "disengaged" from immoral actions, saying that "selective activation and disengagement of internal control permits different types of conduct with the same moral standards" (Bandura, 1991, *p*. 71-72). Eight different moral disengagement processes are identified:

- Moral justification – Killing one so that many can live
- Euphemistic labeling – Saying "enhanced interrogation" instead of torture
- Advantageous comparison – My crimes are minor compares to others'
- Displacement of responsibility – I'm just following orders
- Diffusion of responsibility – Everyone's doing it
- Ignoring or misrepresenting consequences – It's not really that bad
- Dehumanization – They're not really people
- Attribution of blame – It's their fault

Although these processes were mainly analyzed within the context of violent behavior, they should theoretically apply to other types of immoral behavior (Bandura, 1991). Later studies found empirical support for the predictions on moral disengagement (see Bandura, Barbaranelli, Caprara, & Pastoelli, 1996), though the authors noted that the different processes of moral disengagement often overlapped and were difficult to isolate.

Intuitive Perceptive and Synthesis

Whereas the rationalist perspective maintains that moral reasoning and action are based on conscious deliberation, the intuitive perspective posits that many, if not the majority, of our moral judgments are made automatically and unconsciously. This perspective emerged as a response to the rationalist perspective and challenged it on empirical grounds. One of the first studies to test this perspective found that affect was not "post-cognitive" (i.e., only occurring after cognitive processes), but rather was independent of cognition or even memory (Zajonc, 1980). This study determined that affect and cognition were different processes, and though they certainly influenced each other they "constitute independent sources of effects in information processing" (Zajonc, 1980, *p*. 151). Building from this perspective, later studies investigated self-regulation and argued that automatic processes "perform the lion's share of the self-regulatory burden" (Bargh & Chartrand, 1999, *p*. 462). In the realm of moral judgments, it was found that an intuitionist perspective made more accurate predictions than the rationalist model (Haidt, 2001).

Soon after these publications, several studies sought to merge this idea of intuitive judgments with the existing rationalist perspective (Greene & Haidt, 2002; Greene, Nystron, Engel, Darley, & Cohen, 2004). Focused on cognitive neuroscience, these studies proposed that automatic, affective judgments and deliberate, cognitive judgments originated from "competing subsystems in the brain" and as such could, at times, work against each other (Greene et al., 2004, *p*. 389). The findings of these studies complemented previous work on "moral dumbfounding" that identified situations where people will form intuitive moral judgments that they cannot rationally justify (Haidt, Bjorklund, & Murphy, 2000).

Based on these findings, the rationalist and intuitive approaches to morality were synthesized into a dual-process model that argues that most of our moral judgments occur automatically and we only cognitively reason on moral issues when presented with a challenging moral situation. One prominent theory that uses this dual-process model is Moral Foundations Theory (Haidt & Joseph, 2004), which argues that people have five (or potentially more) "moral foundations" that can be found across cultures and govern how we make moral judgments. For example, the authors argue that everyone has a "care/harm" foundation that deals with moral judgments toward helping or hurting others, though what is perceived to be moral behavior along the "care/harm" foundation can vary according to culture and upbringing. The strength of our moral foundations, and the cultural context in which they developed, greatly influence our moral judgments, especially our intuitive ones (Haidt & Graham, 2007). Moral Foundations Theory has been found to reliably predict phenomena such as political orientation (Graham, Haidt, & Nosek, 2009) and attitudes toward criminal behavior such as assault (Charkoff & Young, 2014) and suicide (Rottman, Keleman, & Young, 2014).

Morality and Media

Although there are many theories that look at the relationship between morality and media consumption, I will be considering two: (1) the Moral Disengagement Model (Hartmann & Vorderer, 2010), based on the rationalist perspective, and (2) the Model of Intuitive Morality and Exemplars (Tamborini, 2011), based on the dual-process perspective.

The Moral Disengagement Model applies Bandura's concept of moral disengagement to media enjoyment, specifically violent video games. Noting that violent games include content that violate the moral standards of most people, the model seeks to explain why many people still enjoy playing violent games. The model is based on two assumptions: (1) virtual agents in video games are automatically anthropomorphized by players, and (2) harming those agents would violate (most) players' moral standards, resulting in self-condemnation, guilt, and a less enjoyable experience. As such, the model proposes that many games include "moral disengagement cues," or narrative elements that prompt moral disengagement, thereby mitigating self-condemnation and increasing enjoyment. These cues include dehumanizing enemies by making them literally inhuman (e.g., zombies), distorting the consequences of game violence by portraying violent unrealistically, or justifying violence against enemies through the narrative (Hartmann & Vorderer, 2010).

The Moral Disengagement Model has substantial empirical support. For example, one study found that the presence of moral disengagement cues increased the willingness of players to commit violence (Hartmann, 2012) and a content analysis of popular violent video games found that moral disengagement cues are common in first-person-shooter games (Hartmann, Krakowaik, & Ysay-Vogel, 2014).

In contrast, the Model of Intuitive Morality and Exemplars explains media enjoyment in terms of congruence with the consumer's salient moral foundations. Specifically, we enjoy media that aligns with our salient moral foundation, or that violates them in a way that is easy to categorize, more than we enjoy media violates our salient foundation or that is morally complicated (Tamborini, 2011). Although categorizing whether media aligns with a moral foundation is difficult, both because of cultural differences (Haidt & Joseph, 2004) and because very few narratives load onto only one moral foundation (Clifford, Iyengar, Cabeza, & Sinnott-Armstrong, 2015), this model has also been empirically supported. Within the realm of video games, moral choice within the game is strongly predicted by the saliency of the relevant moral foundations (Joeckel, Bowman, & Dogruel, 2012), and violation of moral foundation within a game can lead to feelings of guilt (Weaver & Lewis, 2012; Grizzard, Tamborini, Lewis, Wang, & Prabhu, 2014).

Hypotheses

To sum up, these two models make different predictions as to what will determine in-game moral choice. The Moral Disengagement Model posits that the presence or absence of moral disengagement cues will predict moral choice, whereas the Model of Intuitive Morality and Exemplars posits that choice will be predicted by the congruence of the player's relevant moral foundations and the moral choice presented.

To test these different predictions, this study presented participants with a simple moral choice: kill or spare a virtual agent. The moral justification associated with killing the agent was manipulated through a narrative cue with a high-justification condition (i.e., the agent was a war criminal), a low-justification condition (i.e., the agent was a petty criminal), or a control condition where no narrative information was given. According to the Moral Disengagement Model, we predict that:

H1: participants will be more likely to kill the virtual agent in the high justification condition than in the low justification condition

H2: killing the agent in the low justification condition will lead to higher levels of guilt than killing the agent in the high justification condition

From the Model of Intuitions and Moral Exemplars, we predict the following:

H3: higher trait levels of the care/harm moral foundation will correlate with a lower likelihood of killing the virtual agent

H4: killing the virtual agent with higher trait levels of the care/harm moral foundation will correlate with stronger feelings of guilt

Methods

Overview: The study was conducted in two sessions. In the first session, participants recruited through the School of Communication participant pool filled out a consent form and complete a survey on Qualtrics. This survey included the Moral Foundations Questionnaire, measures of trait empathy and aggression, self-reports of gaming habits, and demographic information. This survey was administered one week before the second session to avoid priming the participants' behaviors during the experiment.

The second session took place in a laboratory and consisted of the treatment and post-treatment measures. The treatment was a modified version of a video game called *The Killer* (http://www.gametrekking.com/the-games/cambodia/the-killer/play-now), which is a minimalistic side-scroller video game. A side-scroller is a 2-dimensional video game where the player travels linearly through the game environment, typically from left to right. This design was chosen because it reduces potential variance in participant behavior within the game. In this game, the player took the role of an armed stick figure tasked with executing a second stick figure. The player must march the second figure to a designated area and then decide to kill or spare the second figure (here referred to as the "virtual agent"). The game was modified to allow for a narrative manipulation and to record how long participants deliberate before deciding to kill or spare the second figure. Time of deliberation was included to account for a potential ceiling effect (i.e., it may be the case that nearly all of the participants decide to kill the virtual agent, so time to deliberate might be a useful alternative measure, with longer deliberation times corresponding to decreased wiliness to kill the virtual agent).

Immediately after gameplay, participants filled out a measure of guilt, completed a tangram task to measure prosocial/antisocial behavior, completed a manipulation check, and answered an open-ended question. Afterwards, participants were debriefed and informed of the true purpose of the study.

Participants: Participants were 373 undergraduate students (60.3% female, mean age = 20.1, *SD* = 1.81) that were recruited through the Ohio State University School of Communication's participant pool, which is a requirement for introductory courses.

Manipulation: Participants were randomly assigned to one of three conditions: high justification, low justification, or control. The conditions were identical except for a short narrative giving information about the virtual agent at the beginning of the game. In the high justification condition, the participant was told the virtual agent was a "war criminal." In the low justification condition, participants were told the virtual agent was a "petty criminal." In the control condition, no information about the virtual agent was given. The descriptions of the virtual agent were brief, both due to technical limitations, to keep the manipulation simple, and to avoid loading on other mechanisms of moral disengagement, i.e. one of the other processes identified by Bandura.

Measures: Trait aggression was measured using the Aggression Questionnaire (Buss & Perry, 1992), which contains 29 items (e.g. "If somebody hits me, I hit back"; 1 = *Strongly disagree* to 5 = *Strongly agree*, Cronbach α = .89). Trait empathy was measured using an empathy questionnaire (from Raney, 2002), which contains 11 items (e.g. "before criticizing someone, I try to image how I would feel if I were in their place"; 1 = *Strongly disagree* to 5 = *Strongly agree*, Cronbach α = .76). Moral Foundations was measured using the Moral

Foundations Questionnaire (Graham, Nosek, Haidt,, Iyer, Koleva, & Ditto, 2011), which contains 20 items that assess five moral foundations: (1) *harm/care* (whether someone was harmed; Cronbach α =.58), (2) *fairness/reciprocity* (whether everyone was treated equally; Cronbach α= .51), (3) *ingroup/loyalty* (whether the good of the group was taken into account; Cronbach α = .52), (4) *authority*/*respect* (whether authority was respected; Cronbach α= .46), and (5) *purity/sanctity* (whether the situation violated purity; Cronbach α = .51). Items are scored on a scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*.

Guilt was measured by a subset of the Anticipated Guilt, Shame, Pride Scale and Moral Disgust Scale (Marschall, Sanfer, & Tangney, 1994; Nabi, 2002, see appendix for full list of questions used), which contains 22 items (e.g. "I felt remorse, regret"; 1 = *I did not feel this way at all* to 5 = *I felt this way very strongly*, Cronbach α = .65). and also includes questions relevant to video games (e.g. "I knew it was just a game"). Antisocial behavior was measured using a Tangram Task procedure (Saleem, Anderson, & Barlett, 2015), in which participants are told to select puzzles for a fictitious partner. The puzzles are of varying difficulty and the participants are told that their partner will receive $10 if he or she can solve all of the puzzles within 10 minutes. Thus, participants can behave in an antisocial way by assigning their partner many difficult puzzles to solve or in a prosocial way by assigning easy puzzles. The difference score method will be used, where choosing more difficult puzzles vis-à-vis easy puzzles corresponds to higher levels of antisocial behavior and lower levels of prosocial behavior.

Results

Preliminary analysis: 373 participants completed the first session, with 253 also completing the second session. An additional 45 participants were dropped due to incomplete or duplicated entries, leaving 208 participants (58.6% female, mean age = 20.2, *SD* = 2.24). The participants who completed the consent form but did not complete the lab session were not statistically different from the participants who completed both sessions along any of the pre-test measures (see Table 1).

Men were significantly more likely to kill the agent than women (61.5% for men, 41.5% for women, $p < .05$). There were no gender differences in frequency of playing video games ($p = .68$). Males had higher trait aggressiveness scores ($M = 2.68$, $SD = .45$) than females ($M = 2.36$, $SD = .61$), $t(209) = 4.44$, $p < .01$, $d = .57$, see figure 1). In all three conditions, about an equal number of participants chose to kill or spare the agent, and a chi-squared test showed no significant differences between sample sizes across conditions ($p = .59$). The distribution of trait empathy and the moral foundations, separated by gender, are presented in figures 2-7.

The tangram task had a bimodal distribution with one mode at -1 (13.4%) and another at -10 (11.1%), the respective minimum and maximum values for helpfulness (positive numbers, in contrast, represent antisocial or harming behavior). There was no significant correlation between the tangram task and guilt ($p = .30$), between the tangram task and conditions ($p = .98$), or based on whether the participant killed or spared the agent ($p = .87$), though there was a significant, positive correlation between the tangram task and trait aggression ($t = -2.45$, $p < .05$) meaning that more aggressive participants were less helpful. Men were slightly less helpful ($M = -1.33$ for men, -2.29 for women), but the difference was not significant, ($t(189) = 1.49$, $p = .14$, $d = .21$).

Guilt (M = 2.95, SD = 0.39) was slightly right-skewed toward feeling more guilty. Gaming habits were not normally distributed, and a majority of participants (70.7%) played less than 1 hour of video games a day on average[1] (see figure 8). Time spent deliberating (M = 12.82, SD = 7.37) was heavily right-skewed (skewness = 4.70) and as such was transformed with a log transformation (skewness = 0.04) for the analysis.

The free response questions were content-analyzed by two independent coders (kappa = .46, $p < .01$) after the conclusion of data collection. Coders were trained over the course of a week, which consisted of instruction on how to code free-response data and sample data to code, which was reviewed by the researcher until it was accurate. Due to the low kappa, all disagreements were reviewed by the lead researcher, who made the final determination. When asked who the agent was, 79.3% (out of 208 responses) identified the agent as a human of some sort ("a prisoner," "a murderer", i.e. anthropomorphizing the agent), whereas only 5.8% identified the agent as an inhuman entity ("a stick figure," "a bunch of pixels"). The remainder said that they were not sure who the agent was or supplied who they believed their avatar to be instead of the agent. There was no significant correlation between identifying the agent as human and condition, nor any significant correlation between identifying the agent as human and killing/sparing the agent (p = .89). Only 17.8% specified that the agent had committed some sort of crime.

When asked why they killed or spared the agent, 11.3% (out of 197 responses) said they had done so accidentally, 10.7% said they didn't realize they had a choice, 35.0% said they thought they were supposed to ("the game told me to aim," "I thought that was the purpose of the

[1] Recording gaming habits as a dichotomous variable only yielded marginal (insignificant) improvements to models

game"), and 4.5% said that "it was just a game." 32.0% said they spared the agent because they felt that the agent did not deserve to die ("his crimes did not warrant the punishment," "he hadn't done anything to me"). This response was significantly correlated with the low justification condition ($p < .05$), but not the high justification condition ($p = .54$) or the control ($p = .74$). 4.1% said that the agent deserved to die ("he was a rapist," "he had killed too many people"). There was a significant correlation between sparing the agent and reporting that the agent did not deserve to die (phi = .73, $p < .01$), as well as between reporting that they believed they were supposed to kill the agent and killing the agent (phi = .56, $p < .01$).

When asked how they felt about their decision, 21.3% (out of 127 responses) said it was "just a game" and they had no strong feelings one way or another, 39.4% felt good about their decision, and 22.0% reported feeling guilt. There were significant correlations between reporting feeling good about their decision and sparing the agent (phi = .67, $p < .01$), between feeling guilt and killing the agent (phi = .34, $p < .01$), and between feeling that it was "just a game" and killing the agent (phi = .48, $p < .01$). Graphs of the free response answers can be found in figures 9-11 and graphs of deliberation time in figures 12 and 13.

In all analyses involving interactions, continuous predictor variables were centered by subtracting the mean before making the interaction terms.

Predictors of killing the agent: According to H1 (participants will be more likely to kill the virtual agent in the high justification condition than in the low justification condition), the condition the participant is assigned to should predict whether he or she kills or spares the agent, This hypothesis was tested with probit models, presented in in Table 2. Relevant predictors for H1 are trait aggression (colinear with gender), trait empathy, age, gaming habits, and conditions. No interaction terms were found, nor were any of the free response questions significant. Trait

aggression had a significant, positive correlation with killing the agent (B = 0.43, SE = 0.17, p < .05), and gaming habits had a significant, negative correlation with killing the agent (B = -0.18, SE = 0.09, p < 0.5). When added to the model, the free response questions of believing the agent deserved mercy was significantly negatively correlated with killing the agent (B = -2.56, SE = .43, p < .05), and believing that the game has instructed participants to kill the agent (rather than it being a choice) was significantly positively correlated with killing the agent (B = .78, SE = .25, p < .01), though trait aggression and gaming habits were no longer significant in the model. It is important to note that none of the conditions were significant predictors of the likelihood of killing the agent.

However, it may be that using killing or sparing the agent is not granular enough to identify the effects. As such, OLS regression using the natural log of time spent deliberating before killing or sparing the agent was also explored, as shown in Table 3. No significant interaction terms or free response questions were found. A stepwise regression created a model with the low justification condition (B = .12, SE = .08, p = 0.12) and the high justification condition (B = .2, SE = .08, p < .05), however the low justification condition did not significantly contribute to the model when tested ($F(1, 204) = 2.46$, p = .12), and it was dropped. Note that the coefficient for the high justification condition is pointed in the opposite direction than was predicted.

To test H3 (higher trait levels of the care/harm moral foundation will correlate with a lower likelihood of killing the virtual agent), the same kind of models were created, though replacing the conditions with the participants' moral foundations (see Table 4). A stepwise regression (used due to the number of moral foundations and interactions) included trait aggression (B = .43, SE = .17, p < .05), which had a significant positive correlation with killing,

gaming habits (B = -.18, SE = .09, p < .05), which had a significant negative correlation with killing, and trait empathy (B = -.30, SE = .18, p = .09), which had an insignificant negative correlation with killing. No significant interaction terms were found. Once again, adding in the free response questions of thinking the agent deserved mercy (B = -2.58, SE = .44, p < .01) and thinking killing was required (B = .72, SE = .26, p < .01) significantly contributed to the model (chi-squared = 34.3, df = 1, p < .01 and chi-squared = 7.9, df = 1, p < .01, respectively), but all of the other terms became insignificant. None of the moral foundations were significant by themselves. Using deliberation time didn't produce any significant effects.

Predictors of guilt: H2 (killing the agent in the low justification condition will lead to higher levels of guilt than killing the agent in the high justification condition) was tested with OLS regression models, shown in Table 5. No interaction terms were found, nor were the free response questions significant. The only significant correlation was with gaming habits (B = -.07, SE = .02, p < .01), which indicated that people who played video games more often felt less guilty killing the agent.

H4 (killing the virtual agent with higher trait levels of the care/harm moral foundation will correlate with stronger feelings of guilt) was tested in the same manner, shown in Table 6. A stepwise regression identified game habits (B = -.06, SE = .02, p < .05), the fairness foundation (B = .03, SE = .01, p < .01), and the authority foundation (B = -.02, SE = .01, p = .08) as predictors. No interactions were found nor were any free response questions significant.

For each of these four hypotheses, the models were retested after removing participants who reported not understanding what they were doing when killing the agent (n = 23), then further excluding those who did not realize they had a choice (n = 21), then further excluding those who thought they were "supposed to" (n = 69). Removing these participants marginally

improved the AIC of the models but did not change any of the models, with the exception of H2 where gaming habits had a significant negative correlation with guilt (B = -0.06, SE = 0.03, $p < .05$) when the first 23 participants were excluded (further exclusions yielded similar results).

In addition, interactions were tested after centering each of the continuous variables. Once again, no significant interactions were found.

Discussion

This study failed to reject the null hypotheses for all four proposed relations. H1 is rather conclusively unsupported as there was no difference in killing between conditions, nor any interaction with conditions, and the high justification condition was correlated with longer deliberation time rather than shorter. H2 is also unsupported, as the conditions had no significant impact on the models. H3 was entirely unsupported, as was H4 since as fairness and authority, not care, were identified as predictors of guilt. Although the hypotheses were not supported, the study's results do provide some interesting insight.

First, the lack of support for the Moral Disengagement Model seems to be at odds with previous research on the topic. There are several potential explanations for why the predicted effects were not found. The most likely is that the manipulation was simply not strong enough to elicit a powerful response from the participants. Previous studies have made clear distinctions between virtual agents that should or should not elicit moral disengagement, such as human characters versus zombie characters (Hartmann & Vorderer, 2010), or made stark narrative distinctions between conditions, such as liberating prisoners at a torture camp or protecting the camp from liberators (Hartmann, Toz, & Brandon, 2010). In contrast, the stimuli in this study were minimalistic, having identical stick figures for all three conditions and only providing a few lines of narrative to distinguish between conditions. As mentioned above, this was done intentionally, but as a consequence, it is likely that the effect of those stimuli was too small to capture with the number of participants.

In addition, many participants expressed that they either did not know they had a choice in killing or sparing the agent (10.7%), that they believed they were "supposed" to kill the agent

as the objective of the game (35.0%) (note that the game never gave instructions on whether to kill or spare the agent), or they weren't sure what they were doing in the game (11.7%), meaning that over half of participants did not understand that they were able to make a moral choice. The issue of not understanding the option of choice is very likely in part caused by the relative video game illiteracy among the participants. As mentioned above, over 70% of participants played less than 1 hour of video games daily, with a median time among them of 8 minutes per day. Given that the stimulus offered no explicit instructions to make a choice whether to kill or spare the agent, it is likely that those unfamiliar with games in general would not have realized that they could make that choice. However, due to the lack of variation of game habits among the participation, this relationship could not be tested.

Another potential explanation is that the use of stick figures failed to elicit anthropomorphization among the participants, meaning that there would be no reason to morally disengage. However, this doesn't seem to be supported by the free response questions, both as significant predictors and as qualitative data. As mentioned above, only a small minority of participants identified the agent as a non-human entity, and a similar minority expressed feeling no emotional response to killing or sparing the agent. One participant said:

> *I felt no attachment or remorse due the the* [sic] *low quality of graphics and the lack of narrative involvement ssociated* [sic] *with the other character. I have no idea who they are or what they did, or why i should spare them. There was no emotional involvement in this game for me at all.*

This response is typical of the sentiments expressed by this portion of the participants—it's just a game, it didn't mean anything to them. However, a much larger number of participants highly anthropomorphized the agent, some even expressing strong emotions during and after gameplay:

I made the decision to spare the character because I felt gross and uneasy holding a gun to another human being, even in just a stick figure form. I did not want to play God and determine the fate of a human life, so I let the other character go free

---

I decided to spare them because I consider myself a pacifist and also had been given no reason why this person deserved to die. I feel good about my decision, I think; however, throughout the whole game, even before I realized what the game was simulating, I felt unnerved. That feeling of discontent and unease lingered with me after the game was complete.

I got a pit in my stomach when I shot him. I didn't even know there was an option for sparing the character; if I had known that, I would have definitely not killed him. That just feels so wrong.

Most responses fell between these two extremes, though a majority (70%) reported having some sort of emotional response to the game.

These responses suggest two things. First, they seem to rule out a lack of anthropomorphization as an explanation for why we do not see difference in rates of killing between conditions, as most participants imbued human characteristics on the agent. More interestingly, this suggests that the lower limit for when people will anthropomorphize virtual agents is extremely low since literal stick figures were able to elicit these strong emotions out of many participants. This challenges an assumption in previous work on the Moral Disengagement Model that the visual realism of virtual agents is an important factor that distinguishes how people respond to those virtual agents as opposed to objects such as chess pieces (Hartmann,

Toz, & Brandon, 2010). Indeed, many chess sets will include figures that are more realistic than the stick figures used in this study, yet the stick figures still managed to evoke strong anthropomorphization among some participants.

The same issues of weak stimuli and low literacy are likely part of why such anemic results were found in testing the Model of Intuitive Morality and Exemplars as well. However, the fact that care was not significant in any of the models, but other moral foundations were, is both interesting and difficult to explain. Part of the explanation is likely the low inter-item consistency between many of the items on the Moral Foundations Questionnaire used (the reported Cronbach's alphas here are far lower than the typical measure using the Moral Foundations Questionnaire), meaning that the survey may have not done a good job in capturing the moral foundations of the participants. However, I do not have a solid explanation as to why the Questionnaire underperformed in this case. One potential explanation is that there were many participants in this study who did not speak English as their native language (though all reported being "fluent" in English), but I do not have the data to parse between native and non-native speakers within the sample. Another potential explanation is the problem encountered by Clifford et al. (2015)—it is nearly impossible to construct situations that load onto only one moral foundation, and the stimuli used in this study seem to be no exception.

Looking at the guilt model, the moral foundations identified do point in the expected directions. We would expect high levels of the fairness foundation to increase guilt, as the situation between the participants and the agent were very unfair. Similarly, we would expect high levels of the authority foundation to decrease guilt as the game has an implied source of authority (whoever tasked the participant with executing the agent), or the participants might see the game itself as a source of authority. These results highlight the need to further examine how

to create situations which load onto the desired moral foundation and suggest that seemingly straightforward moral dilemmas—kill or spare a character—can invoke very complicated moral intuitions.

This study has several interesting implications, however the results found are so weak that future research that uses a stronger manipulation and better measurements on the relevant variables will be needed to draw definitive conclusions. However, it does seem apparent from this study that very simple representations of people and situations can lead to anthropomorphization and evoke strong emotions. Although this finding should not be terribly surprising considering that participants have been shown to anthropomorphize triangles and circles (Heider & Simmel, 1944), it does suggest that current research may place too much emphasis on the importance of how realistic virtual agents are.

Future research in this area should again try to compare the prediction of the Moral Disengagement Model and the Model of Intuitive Morality and Exemplars, implementing the suggestions mentioned above. It would also be beneficial to present a variety of moral dilemmas rather than just one to see if the effects of the models are consistent across different types of moral issues. In addition, prior attitudes toward those moral issues should be measured in a pre-test survey. For example, this study would have greatly benefitted from measuring the participants' attitudes toward the death penalty, and the lack of those data is a limitation of the findings. Future research should also continue to investigate the limits, if any, there are to anthropomorphization. It may be the case that some explicit or implied narrative matters much more than the realism of characters. In addition, this study did not measure game skills, but rather gaming habits. There may be an important difference between these two concepts, and measuring gaming skill, or at least self-reported gaming skill, would be beneficial. Finally, a

prompt to remember the text of the manipulation might have improved the manipulation's efficacy—for example instructing the participants that they will be tested on who the virtual agent was at the end of the study. It may be the case that the lower power of the manipulation was in part due to participants not paying close attention to it in the first place.

As this study highlights, the empirical study of morality is notoriously difficult. Even relatively simple questions such as "why do people enjoy violent stories?" open up a Pandora's Box of difficult to define terms, moral concepts which are nearly impossible to isolate, difficulties in operationalizations, and several competing theories that explain the same phenomena. This is an area of research that will be difficult to untangle but which promises to elucidate on a fundamental aspect of human behavior. I hope that this study has been able to make some small contribution in that effort and that future research can build off of the findings and limitations identified here.

References

Bandura, A. (1991). Social Cognitive Theory of moral thought and action. In W. Kurtines & J. Gewirtz (eds.), *Handbook of Moral Behavior and Development* (45-103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, *71*(2), 364-374.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462-479.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology* 63(3), 452-459.

Chakroff, A., & Young, L. (2015). Harmful situations, impure people: An attribution asymmetry across moral domains. *Cognition*, *136*, 30-37.

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, *47*(4), 1178-1198.

Graham, J., Haidt, J., & Nosek, *B.* A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029-1046.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366-385.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389-400.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517-523.

Grizzard, M., Tamborini, R., Lewis, R. J., Wang, L., & Prabhu, S. (2014). Being bad in a video game can make us morally sensitive. *Cyberpsychology, Behavior, and Social Networking, 17*(8), 499-504.

Haidt, J. (2001).. *Psychological Review*, *108*(4), 814-834.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98-116.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*(4), 55-66.

Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Unpublished manuscript,* University of Virginia.

Hartmann, T. (2012). Moral disengagement during exposure to media violence: Would it feel right to shoot an innocent civilian in a video game?. In *Media and the Moral Mind* (pp. 133-155). Routledge.

Hartmann, T., & Vorderer, P. (2010). It's okay to shoot a character: Moral disengagement in violent video games. *Journal of Communication*, *60*(1), 94-119.

Hartmann, T., Krakowiak, K. M., & Tsay-Vogel, M. (2014). How violent video games communicate violence: A literature review and content analysis of moral disengagement factors. *Communication Monographs*, *81*(3), 310-332.

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, *13*(4), 339-363.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243-259.

Joeckel, S., Bowman, N. D., & Dogruel, L. (2012). Gut or game? The influence of moral intuitions on decisions in video games. *Media Psychology*, *15*(4), 460-485.

Kohlberg, L. (1971). Stages of moral development. *Moral Education*, *1*, 23-92.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-development approach. *Moral Development and Behavior: Theory Research and Social Issues,* 31-53.

Kohlberg, L., Levine, C., & Hewer, A. (1983). *Moral stages: A current formulation and a response to critics*. New York: Karger.

Marschall, D., Sanftner, J., & Tangney, J. *P*. (1994). *The state shame and guilt scale*. Fairfax, VA: George Mason University.

Nabi, R. L. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition & Emotion*, *16*(5), 695-703.

Raney, A. A. (2002). Moral judgment as a predictor of enjoyment of crime drama. *Media Psychology*, *4*, 305-322.

Rottman, J., Kelemen, D., & Young, L. (2014). Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition*, *130*(2), 217-226.

Saleem, M., Anderson, C. A., & Barlett, C. *P*. (2015). Assessing helping and hurting behaviors through the tangram help/hurt task. *Personality and Social Psychology Bulletin*, *41*(10), 1345-1362.

Tamborini, R. (2011). Moral intuition and media entertainment. *Journal of Media Psychology*, *23*(1), 39-45.

Weaver, A. J., & Lewis, N. (2012). Mirrored morality: An exploration of moral choice in video games. *Cyberpsychology, Behavior, and Social Networking*, *15*(11), 610-614.

Zajonc, R. *B.* (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151-175.

Appendix: Tables, Figures, and Questions

Table 1. Mean scores for those who completed only the pretest session and those who completed both sessions on all the pretest variables.

	Pretest only	**Both sessions**	**Test statistic**	**p**
Age	20.01	20.02	t(415) = -0.44	.66
Gender (1 = female)	.62	.59	Chisq(1) = 0.62	.43
Trait aggression	2.49	2.47	t(456) = -0.39	.69
Trait empathy	3.65	3.64	t(402) = -0.29	.77
Care foundation	16.03	16	t(431) = -0.15	.88
Fairness foundation	15.6	15.55	t(417) = -0.29	.77
Loyalty foundation	13.57	13.68	t(420) = 0.44	.66
Authority foundation	13.59	13.57	t(440) = -0.07	.94
Purity foundation	13.56	13.56	t(426) = 0.01	.99
Hours of games per day	0.86	0.84	t(449) = -0.18	.85
Number of years playing games	5.38	5.78	t(416) = 0.70	.49

Figure 1.

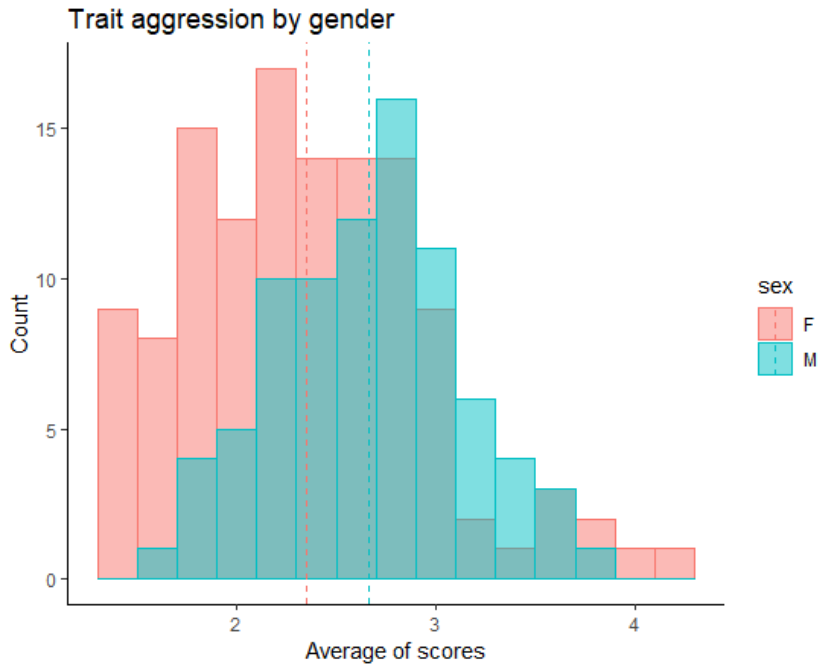

Figure 2.

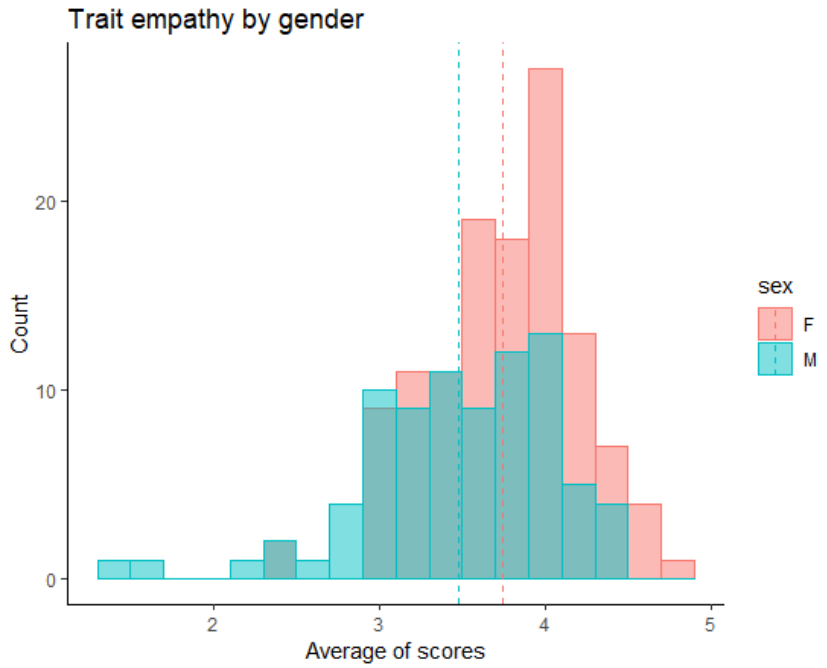

Figure 3.

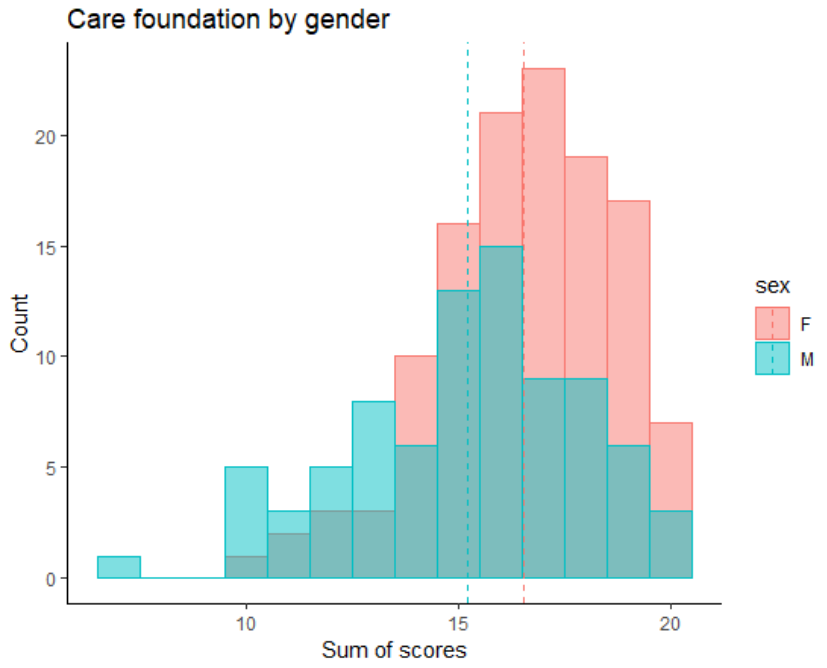

Figure 4.

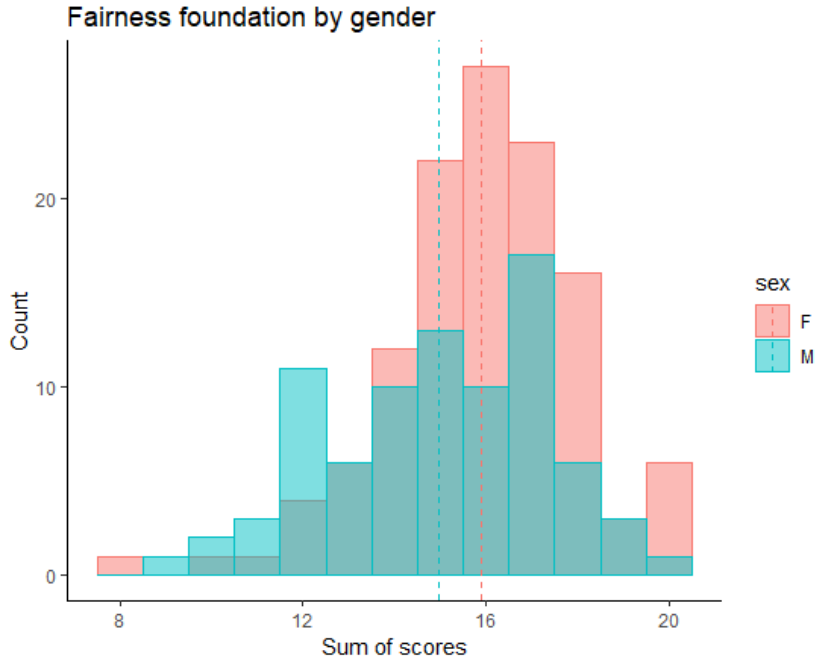

Figure 5.

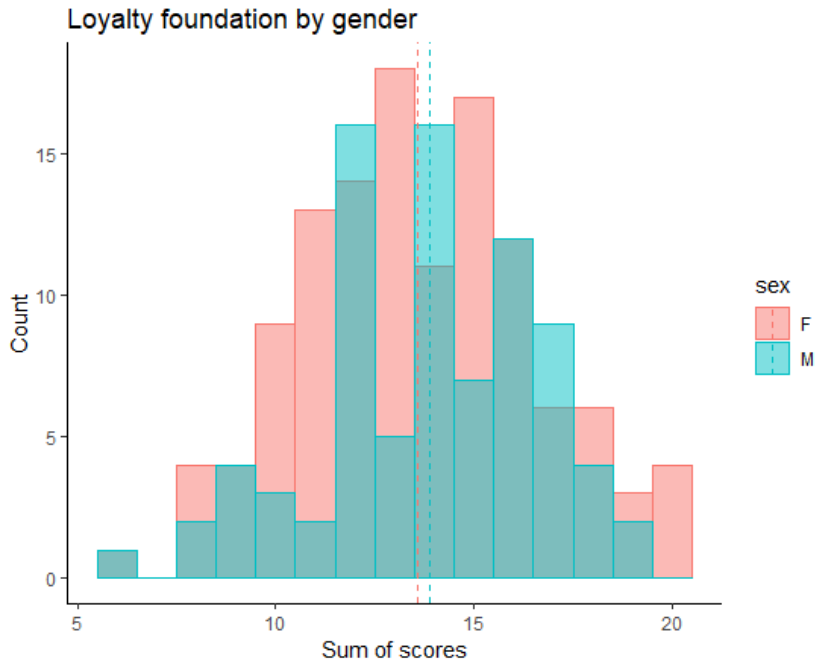

Figure 6.

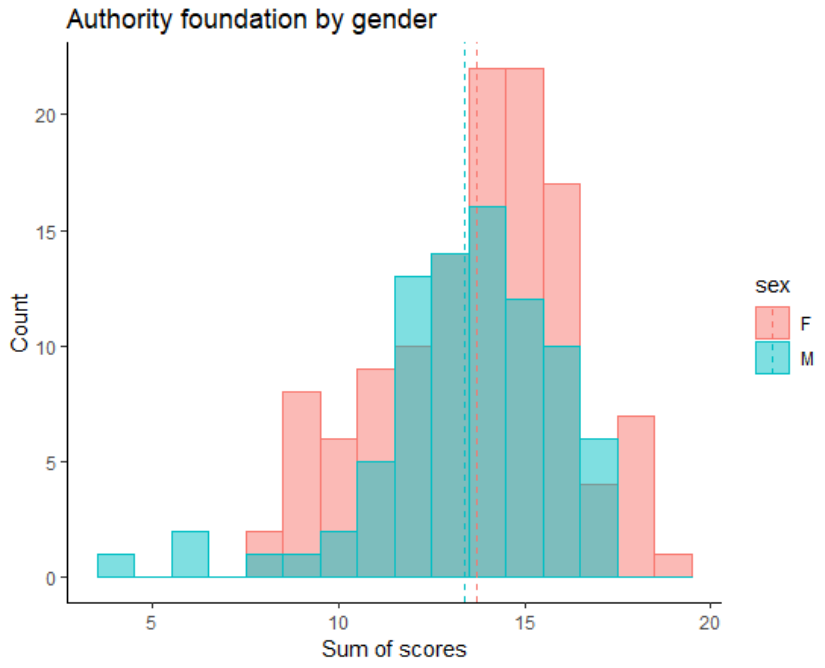

Figure 7.

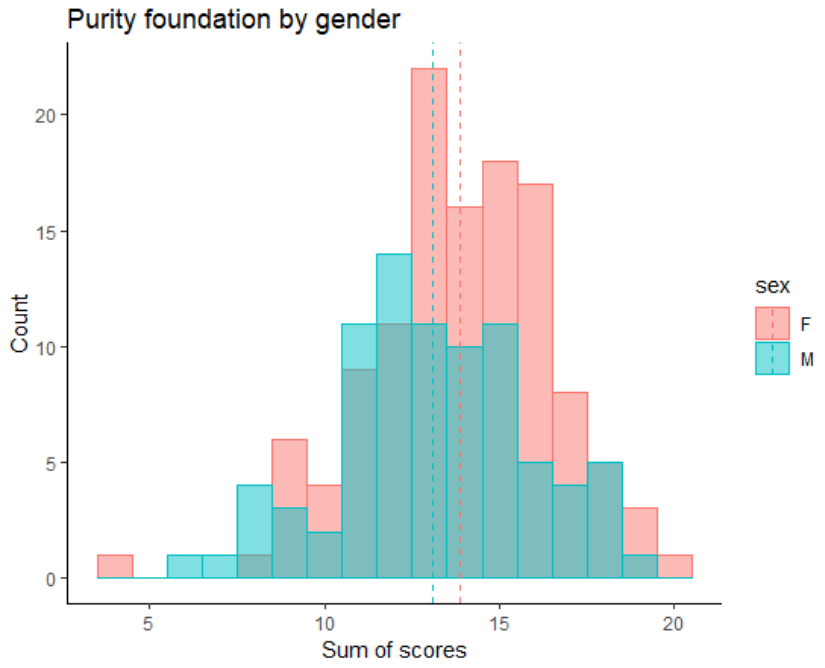

Figure 8.

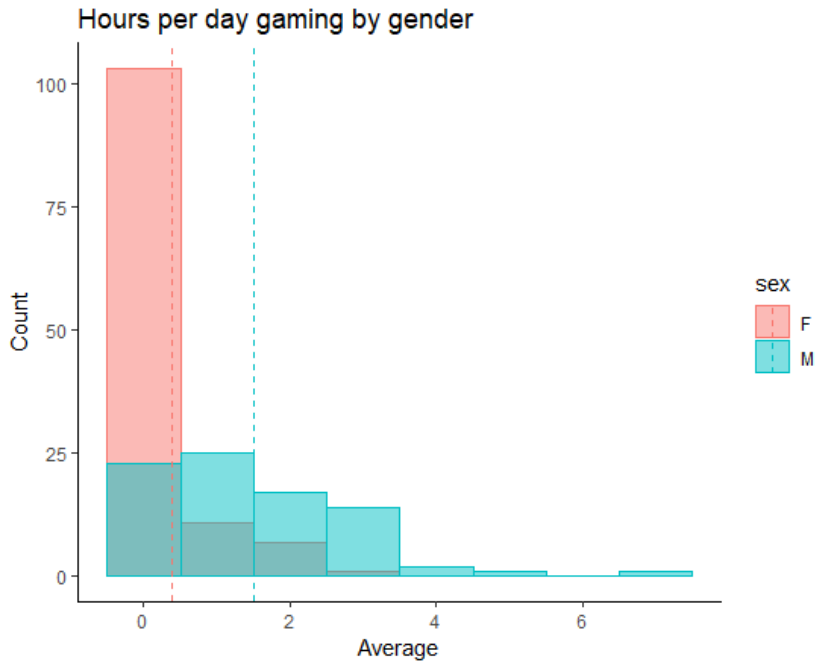

Figure 9.

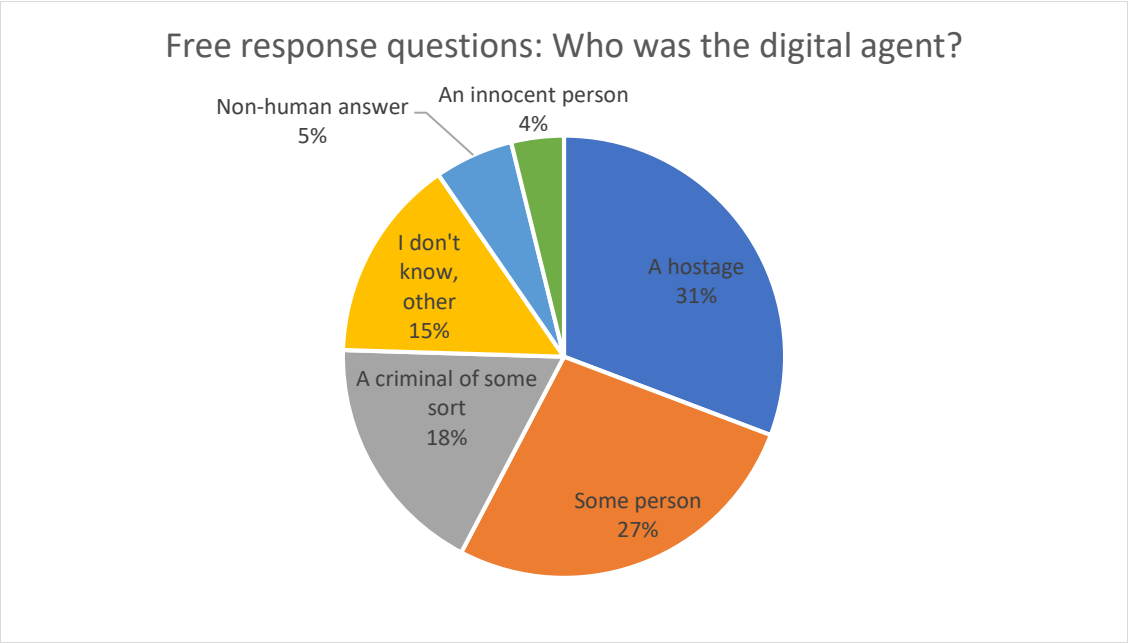

Figure 10.

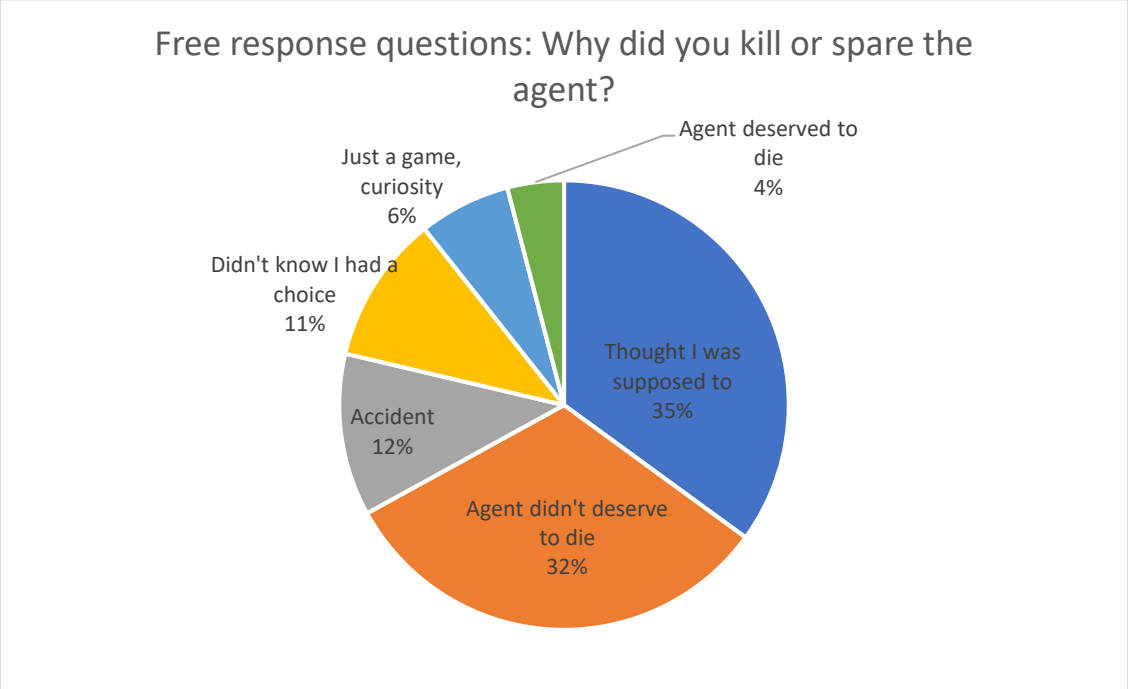

Figure 11.

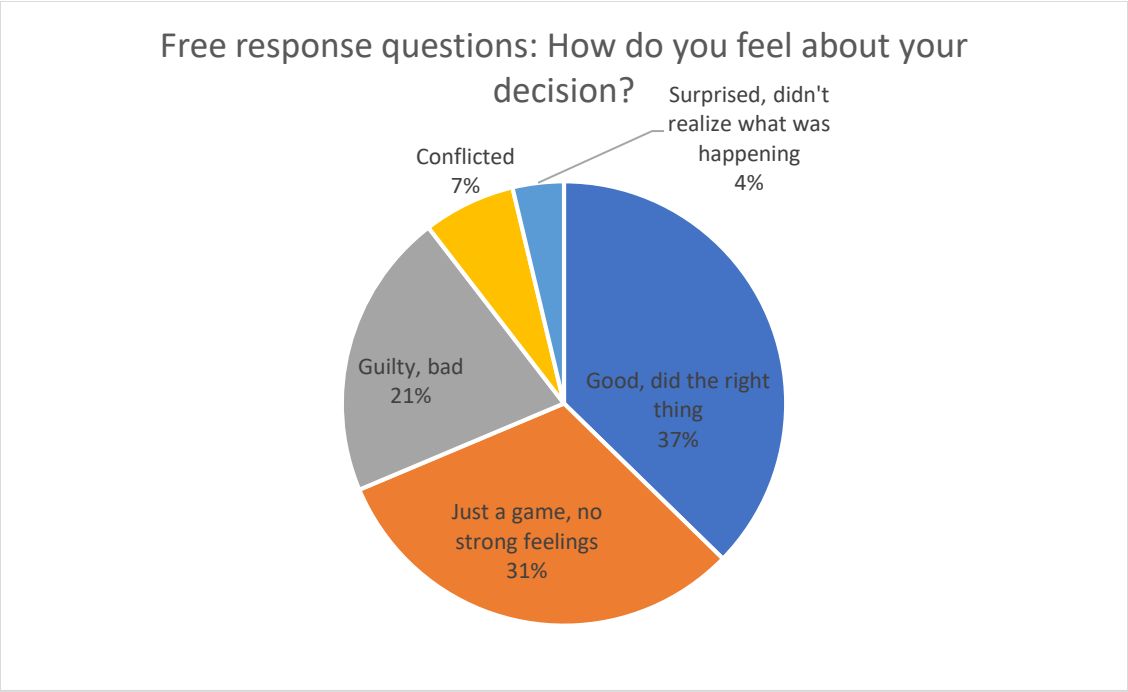
Free response questions: How do you feel about your decision?
Surprised, didn't realize what was happening 4%
Conflicted 7%
Guilty, bad 21%
Good, did the right thing 37%
Just a game, no strong feelings 31%

Figure 12.

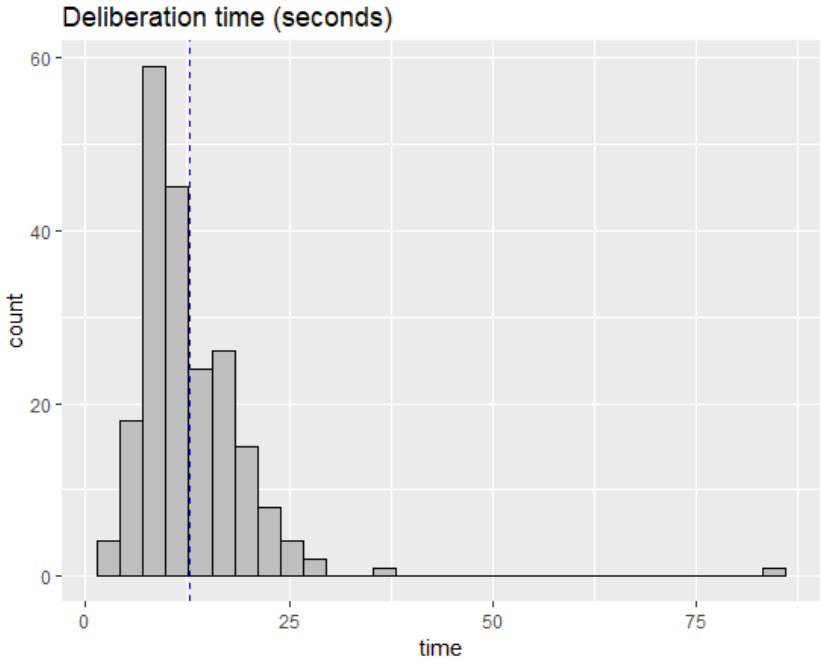

Figure 13.

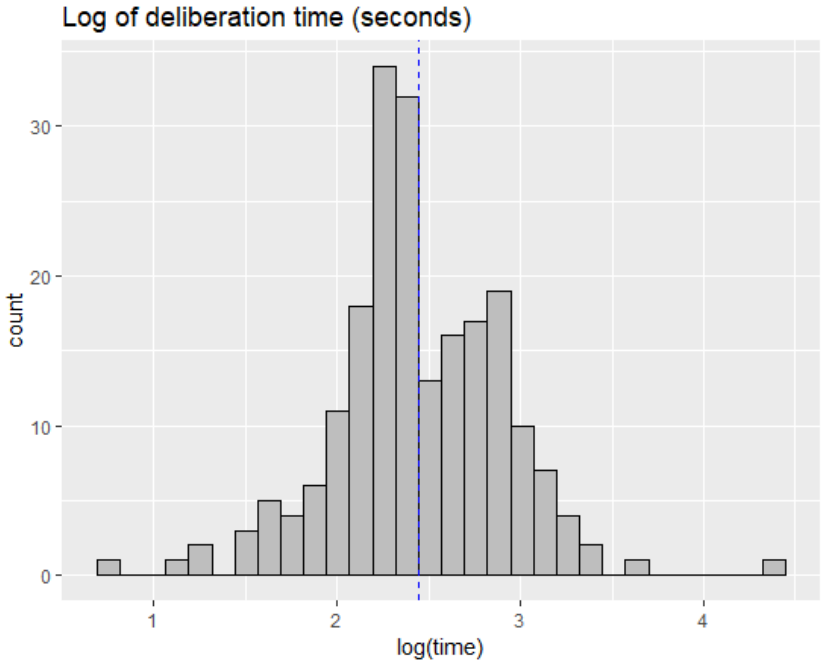

Table 2.

Moral Disengagement Predicting Killing, probit regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Trait aggression	0.077	0.434***								0.494***	0.133
	(0.234)	(0.159)								(0.164)	(0.223)
Trait empathy	-0.264		-0.372**								
	(0.261)		(0.171)								
Age	-0.023			-0.039							
	(0.037)			(0.031)							
High justification	0.389				0.108						
	(0.332)				(0.184)						
Low justification	-0.188					0.085					
	(0.304)					(0.186)					
Gaming habits	-0.066						-0.115			-0.163*	-0.091
	(0.133)						(0.083)			(0.086)	(0.124)
Agent deserved mercy	-2.617***							-2.958***			-2.572***
	(0.451)							(0.415)			(0.439)
Following instructions	0.813***								1.668***		0.731***
	(0.271)								(0.221)		(0.256)
Constant	1.649	-1.030**	1.398**	0.818	-0.000	0.009	0.139	0.810***	-0.470***	-1.037**	0.191
	(1.429)	(0.402)	(0.631)	(0.621)	(0.107)	(0.106)	(0.112)	(0.122)	(0.120)	(0.406)	(0.591)
Akaike Inf. Crit.	148.081	283.117	285.925	290.469	291.834	291.969	288.822	151.639	208.864	281.468	144.814

Note: coefficients (std errors) *p <.1, **p < .05, ***p <.01

Table 3.

Moral Disengagement Predicting Killing, OLS log(deliberation time)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Trait aggression	-0.018	-0.004					
	(0.060)	(0.057)					
Trait empathy	-0.067		-0.054				
	(0.064)		(0.061)				
Age	-0.002			-0.001			
	(0.011)			(0.011)			
High justification	0.204**				0.139**		
	(0.079)				(0.068)		
Low justification	0.123					0.023	
	(0.079)					(0.069)	
Game habits	-0.0001						-0.010
	(0.032)						(0.030)
Constant	2.665***	2.451***	2.639***	2.464***	2.394***	2.434***	2.450***
	(0.374)	(0.145)	(0.226)	(0.221)	(0.039)	(0.039)	(0.041)
R^2	0.038	0.00002	0.004	0.0001	0.020	0.001	0.001
Adjusted R^2	0.009	-0.005	-0.001	-0.005	0.015	-0.004	-0.004
Residual Std. Error	0.462	0.465	0.465	0.465	0.461	0.465	0.465
F Statistic	1.299	0.005	0.786	0.011	4.240**	0.113	0.106

Note: coefficients (std errors) *p <.1, **p < .05, ***p <.01

Table 4.

Moral Foundations Predicting Killing, probit model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Trait aggression	0.082	0.434***							0.427**	0.072
	(0.232)	(0.159)							(0.169)	(0.230)
Trait empathy	-0.325		-0.372**						-0.303*	-0.236
	(0.290)		(0.171)						(0.179)	(0.255)
Care foundation	0.040			0.015						
	(0.060)			(0.035)						
Age	-0.021				-0.039					
	(0.037)				(0.031)					
Game habits	-0.086					-0.115			-0.179**	-0.109
	(0.129)					(0.083)			(0.087)	(0.128)
Agent deserved mercy	-2.513***						-2.958***			-2.578***
	(0.435)						(0.415)			(0.440)
Following instructions	0.763***							1.668***		0.723***
	(0.262)							(0.221)		(0.258)
Constant	1.254	-1.030**	1.398**	-0.198	0.818	0.139	0.810***	-0.470***	0.242	1.221
	(1.482)	(0.402)	(0.631)	(0.572)	(0.621)	(0.112)	(0.122)	(0.120)	(0.853)	(1.236)
Akaike Inf. Crit.	149.097	283.117	285.925	290.546	290.469	288.822	151.639	208.864	280.546	145.986

Note: *p <.1, **p < .05, ***p <.01

coefficients

(std errors)

Table 5.

Moral Disengagement Predicting Guilt, OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Killed agent	0.022	0.007						
	(0.056)	(0.054)						
Trait aggression	-0.072		-0.097**					
	(0.050)		(0.048)					
Trait empathy	0.053			0.085*				
	(0.053)			(0.051)				
Age	0.008				0.003			
	(0.009)				(0.009)			
High justification	0.057					0.076		
	(0.066)					(0.057)		
Low justification	0.008						-0.023	
	(0.066)						(0.058)	
Game habits	-0.056**							-0.065***
	(0.026)							(0.025)
Constant	2.786***	2.948***	3.193***	2.642***	2.884***	2.926***	2.959***	3.007***
	(0.315)	(0.039)	(0.121)	(0.188)	(0.184)	(0.033)	(0.033)	(0.034)
Akaike Inf. Crit.	201.857	202.244	198.108	199.540	202.121	200.453	202.100	195.572

Note: coefficients (std errors) *p <.1, **p < .05, ***p <.01

Table 6.

Moral Foundations Predicting Guilt, OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Killed agent	0.019	0.007								
	(0.055)	(0.054)								
Trait aggression	-0.069		-0.097**							
	(0.050)		(0.048)							
Trait empathy	0.002			0.085*						
	(0.063)			(0.051)						
Age	0.008				0.003					
	(0.009)				(0.009)					
Care foundation	-0.004					0.018				
	(0.014)					(0.011)				
Fairness foundation	0.034**						0.030**			0.033***
	(0.015)						(0.012)			(0.012)
Authority foundation	-0.019*							-0.012		-0.019*
	(0.011)							(0.011)		(0.011)
Game habits	-0.060**								-0.065***	-0.062**
	(0.026)								(0.025)	(0.025)
Constant	2.779***	2.948***	3.193***	2.642***	2.884***	2.664***	2.492***	3.120***	3.007***	2.753***
	(0.344)	(0.039)	(0.121)	(0.188)	(0.184)	(0.177)	(0.186)	(0.151)	(0.034)	(0.212)
Akaike Inf. Crit.	197.583	202.244	198.108	199.540	202.121	199.590	196.067	201.022	195.572	190.544

Note: coefficients (std errors) *p <.1, **p < .05, ***p <.01

Guilt: Adapted anticipated guilt, shame, pride scale and moral disgust scale

Marschall, D., Sanftner, J., & Tangney, J. P. (1994). The state shame and guilt scale. Fairfax,

VA: George Mason University.

Below you find a list of words that you can use to show how you felt while playing the game. Each question asks you about a different feeling. We would like you to tell us how often you felt each of these feelings while playing the game. You can tell us how strongly you felt each of the feelings on the list by marking from 1 (I did not feel this way at all) to 5 (I felt this very strongly)

While playing the game, how often did you:

- I felt remorse, regret
- I felt tension about what I did
- I couldn't help thinking that I did something bad
- I felt like apologizing, confessing
- I felt bad about what I did
- I wanted to sink into the floor and disappear
- I felt small
- I felt like I am a bad person
- I felt worthless, powerless
- I felt humiliated, disgraced
- I felt good about myself
- I felt worthwhile, valuable
- I felt capable, useful
- I felt proud
- I felt pleased about what I did in the game
- I felt cool
- I felt okay, neither particularly happy nor sad
- I knew it was "just a game"
- I felt effective
- I felt grossed out
- I felt disgusted
- I felt repulsed