Imagining wrong: Fictitious contexts mitigate condemnation of harm more than impurity

John S. Sabo and Roger Giner-Sorolla

University of Kent, Canterbury, United Kingdom

Final accepted version, Journal of Experimental Psychology: General

17 October 2016

Author Note

John S. Sabo, Department of Psychology, University of Kent, United Kingdom; Roger S. Giner-Sorolla, Department of Psychology, University of Kent, United Kingdom.

Correspondence concerning this article should be addressed to John Sabo, Department of Psychology, University of Kent, Canterbury, UK, CT2 7NP. E-mail: js766@kent.ac.uk

Word count 11,009

Abstract

Over five experiments we test the fictive pass asymmetry hypothesis. Following observations of ethics and public reactions to media, we propose that fictional contexts, such as reality, imagination, and virtual environments, will mitigate people's moral condemnation of harm violations, more so than purity violations. That is, imagining a purely harmful act is given a "fictive pass," in moral judgment, whereas imagining an abnormal act involving the body is evaluated more negatively because it is seen as more diagnostic of bad character. For Experiment 1, an undergraduate sample (N = 250) evaluated nine vignettes depicting an agent committing either violations of harm or purity in real life, watching them in films, or imagining them. For Experiments 2 and 3, online participants (N = 375 and N = 321, respectively) evaluated a single vignette depicting an agent committing a violation of harm or purity that either occurred in real life, was imagined, watched in a film, or performed in a video game. Experiment 4 (N = 348) used an analysis of moderated mediation to demonstrate that the perceived wrongness of fictional purity violations is explained both by the extent to which they are seen as a cue to, and a cause of, a poor moral character. Lastly, Experiment 5 (N = 484) validated our manipulations and included the presumption of desire as an additional mediator of the fictive pass asymmetry effects. We discuss implications for moral theories of act and character, anger and disgust, and for media use and regulation.

Keywords: moral emotions, moral judgments, moral domains, media, fiction

Imagining wrong: Fictitious contexts mitigate condemnation of harm more than impurity

> "Until I went online and checked the content of this game, I thought it was just a bit of swearing and some shooting and I think some of the parents will tell you that they have been equally naïve." (Iwan, 2014)
>
> – A schoolteacher on the sexual content of 2013's Grand Theft Auto V

Do people have a double standard when it comes to imaginary representations of sex and violence? A number of contemporary video games such as Manhunt 2, No More Heroes, Postal 2, and Reservoir Dogs contain virtual representations of dismemberment, torture, and other graphic acts of violence (Young & Whitty, 2011). Such games, however, did not draw as much controversy as the 2005 Microsoft Windows version of Grand Theft Auto: San Andreas. Although this game involves the player in drug dealing and violence, these were not the reasons that it was eventually withdrawn from the shelves of most major retailers. Rather, the scandal was caused by a mini-game, deleted by the developers but dug up from the video game's code by the gaming community, that allowed one's avatar to engage in fully clothed, poorly animated, and entirely consensual sex.

This tendency to judge fictional sex more harshly than fictional violence has been remarked upon in the context of computer games (Brown v. Entertainment Merchants Association, 2011), and demonstrated in the case of film ratings (Leone, 2004; Thompson & Yokota, 2004; Olson, 2014). The issue has also been breached in ethical philosophy; Luck (2009), for example, acknowledges the existence of this asymmetry in lay morality while making a prescriptive argument against it.

In the present paper, we aim to provide systematic evidence for this phenomenon, which we label the fictive pass asymmetry, and to explain its corollary judgments and emotions. We compare moral judgments of real life acts to moral judgments of imagining those acts, or consuming them in fiction. Normally, judgments of things that are "make-believe" should be less severe – they should get a "pass" in moral judgment. However, we think that this pass is more strongly given to fictions in which immoral acts harm other people, compared to fictions in which immoral acts are not immediately harmful to others but violate norms of purity (that is, involving a counter-normative use of the body, such as violating a sexual taboo). We investigate this hypothesis across various fictive contexts, including media products and one's own imagination.

Recent research has argued that violations of harm and purity, among others, form distinct moral domains (Chakroff, Dungan, & Young, 2013; Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011; Rozin, et al., 1999; Russell, Piazza, & Giner-Sorolla, 2013). Violations of the harm domain are physical acts of violence or deprivation with specific negative outcomes for others. Violations of purity are most centrally those acts that violate bodily moral norms (Giner-Sorolla, Bosson, Caswell, & Hettinger, 2013; Russell & Giner-Sorolla, 2013). This may include sexual acts that are seen as wrong in and of themselves regardless of consent (e.g. incest), food consumption that goes against religious tenets, or body modifications that are seen as a defilement of the self. An alternate view characterizes both types of violation as involving harm, but to different targets; that is, acts classified in the "harm" domain involve injury to specific individuals (as in the original statement of dyadic morality theory; Gray & Wegner, 2010) whereas acts classified in the "purity" domain are seen as harming non-specific entities such as nature or society (as in revised views of that theory; Gray, Schein, & Ward, 2014; see

also Gutierrez & Giner-Sorolla, 2011). In this research, as in much research on purity violations, we focus on abnormal sexual acts that are seen as immoral, which can exemplify either purity violations or impersonal harm; our final study further tests the possibility that our effect, expressed in terms of harm and purity violations, can coexist with a distinction between personal and impersonal harm. Of course, many conceivable acts violate both kinds of norms; however, to test the fictive pass asymmetry hypothesis, we focus on acts that offer explicit demarcation between the two. For convenience, in line with the evolution of our thought throughout this research, and in line with much existing literature we use the terms "harm" and "purity" throughout for the two kinds of violations. The characterization of purity violations as harming impersonal entities was only raised in the final study, and we bring that terminology to bear there.

We were primarily interested in the extent to which negative moral judgments and emotions aroused by judging harm and purity violations would cross the line from real to fantasy contexts. Recent research gives a theoretical context to the fictive pass asymmetry hypothesis by identifying unique ways in which different kinds of immoral acts are evaluated. Firstly, researchers have identified an act-character divide in moral judgments (Pizarro, Tannenbaum, & Uhlmann, 2012; Tannenbaum, Uhlmann, & Diermeier, 2011; Uhlmann, Zhu, & Diermeier, 2014). Acts that directly harm people are seen as bad due to their consequences, while other acts (e. g. violence towards a cat) are seen as more solely indicative of bad character (Tannenbaum et al., 2011). Following this line of thought, Chakroff and Young (2015) demonstrated an attributional asymmetry between the moral domains of harm and purity: impure acts that do not involve harming other people, relative to other-harmful acts that do not involve impurity, are more condemned because of character-based attributions. By contrast, people endorsed more act-

based explanations for acts that harmed individuals. Even though acts experienced through fiction do not harm specific persons, they signal the status of one's moral character, or can even be seen to corrupt and undermine character. Acts that harm other people may be condemned more in real life than immoral acts that do not, because of their consequences for others, but in fiction both acts indicate an engagement with the idea of breaking social norms, that would lead to more equivalent appraisals of bad character.

The moral emotions of anger and disgust are most commonly associated with violations of harm and purity, respectively, although they often co-occur (Russell, Piazza, & Giner-Sorolla, 2013). As in previous research, we expect that real-life purity violations, compared to harm violations, will evoke more disgust relative to anger and vice-versa. We will also explore the possibility that when anger and disgust are distinguishable from one another, fictive pass asymmetry effects will present themselves differently between these emotions. Anger, more so then disgust, is a flexible emotion (Russell & Giner-Sorolla, 2011a; 2011b) associated with harmful behavior. For instance, Piazza, Russell, & Sousa, Study 1 (2013) found that anger, independent of disgust, negatively predicts the envisaging of mitigating circumstances of both harm and purity code violations. Meanwhile disgust, when anger was controlled for, had no relation to participant's envisaging of mitigating circumstances. Consistent with this, Russell and Giner-Sorolla (2011b) instructed participants to justify both harmful and impure acts. The amount of anger that the participants experienced changed as a result of this exercise. Their levels of disgust, on the other hand, were relatively more stable. For the present research, we believe that when anger is directed at fictional versus real behavior, if already at high levels, it should drop more than disgust does. In other words, when one imagines harm to specific others in a fictional context, the amount of anger should diminish, leaving the co-occurrent emotion of

disgust as the more prevalent emotional reaction. This reasoning suggests that the asymmetry should be more evident for measures of anger than for measures of disgust.

To sum up, our fictive pass asymmetry hypothesis predicts that fictitious harm to others will be given a pass and subjected to less condemnation than its real-life counterpart. On the other hand, fictitious moral violations that do not harm specific others ("purity") should be denied a pass. In the strongest expression of this, fictional acts that violate purity norms should be subjected to as much condemnation as their real-life counterparts. However, it could also be that fictional purity violations are condemned less than real-life ones in absolute terms, but that this drop is smaller than the drop between reality and fiction for harm violations.

Across five experiments we examine the effect of a number of different fictional contexts on moral judgments of described acts that violate the moral norms of harm and purity. In each experiment, we compare the context of real life to a number of fictional contexts, such as imagination, film and video games. Since we did not have any specific hypothesis regarding the relative strength of fictional contexts, within each experiment they were collapsed into a single index to facilitate clear reporting. After the last experiment, we also report the outcome of a meta-analysis across a set of comparable experiments that included the same set of fictional contexts, looking at differences in fictive pass effects between the contexts.

To give an overview, Experiment 1 is an initial test of the fictive pass asymmetry hypothesis. Experiment 2 replicates this finding with different media contexts and moral violation examples. Experiments 3, 4 and 5 extend the former experiments by including all of the previously examined fictitious contexts. They sequentially add measures that allowed the design

to address what kind of character judgments mediate the effect, how the effect extends into a desire to punish the offender, and to what extent fictional acts indicate one's true desires.

Methodological Notes

With the exception of Experiment 1 (a university student sample), all participants were recruited using Amazon Mechanical Turk (www.mturk.com). Participants recruited via Mechanical Turk were all living in the USA, and were given financial compensation for their participation. Criteria for excluding participants from the analysis were: unfinished questionnaires, rote responses (e.g. endorsing the same scale point across all items), or failing attention-checking questions. When possible, new participants were recruited in their stead. Sample sizes were determined a priori and not altered based on results. We report all measures and manipulations in these experiments.

Experiment 1

Method

Participants

Two hundred fifty undergraduate psychology students (196 female; M_{age} = 20.3; SD_{age} = 4.1) from a university in southern England participated for course credit. Twenty-six participants (10%) were excluded based on the stated criteria. This sample size, based on availability, was conservatively high for the within-subjects design (i.e., it had over 95% power to detect a small f of 0.15 given r = 0.30 among measures), but we thought this appropriate for an initial test of the hypothesis.

Design

Each participant evaluated nine vignettes that depicted fictional agents committing different types of abnormal behavior (harm, purity, and pathogen, as described below). Within these nine vignettes, three were described as occurring in real life, three as imagined by the main agent, and three as watched in a film by the main agent. Counterbalancing, using 27 different sets of pairings between participants, ensured that each participant responded to nine combinations of the two factors' three levels. The key design for analysis was within-participants, 3 (code: harm vs purity vs pathogen) x 2 (context: reality vs fiction). As previously explained, the two fictional contexts were collapsed to form a single level.

Pretest

The nine vignettes for Experiment 1 were derived from a pretest in which 31 undergraduate participants rated their judgments of levels of anger, disgust, and moral wrongness for twenty vignettes that depicted violations of harm (e.g. falsifying information on a CV to get unjustly hired) and purity (e.g. consensual sibling incest), presented as occurring in real life. In each vignette, one person – the agent – was named and identified as committing the violation. For each code violation, we selected three vignettes that were similar in moral wrongness. Our selections were also confirmed by each selected vignette showing the expected pattern of emotional responses (i.e. anger significantly > disgust for harm and disgust significantly > anger for purity). This selection method allowed us to be confident that our effects would be due to the code violation portrayed by each set of vignettes rather than differences in overall disapproval.

Through pretesting, we also identified vignettes of pathogen violations – that is, physically disgusting, but relatively less morally objectionable violations that threaten personal health (e.g. eating spoiled food), which proved to be high in disgust but relatively low in both

anger and moral violation rating. These were included in this first experiment to test whether pathogen would show a similar lack of fictive pass as predicted for purity violations. All vignettes, contextualized in reality, are reported in the supplementary material, document A.

Materials

The wording of the pretested vignettes were manipulated to depict the acts as occurring in reality or as in a fictional context. For example, in the context of reality a harm violation appeared as "Janice decided to put some false information on her CV in order to make it more impressive. By doing this she managed to get hired over candidates who were actually more qualified". In the context of imagination, the vignette appeared as, "Janice imagined putting some false information on her CV in order to make it more impressive. She imagines that by *doing this she manages to get hired over candidates who were actually more qualified"*.

A one-item measure assessed the moral wrongness of the main agent's act (How morally wrong was [agent*'s name*] *'s* behavior?) from 1 (Not at all) to 7 (Very much). Two items more directly tapped the moral character of the main agent: one item asked "Do you think [agent*'s* name] is mainly a good or bad person?" (1 = mainly good; 7 = mainly bad) and the other asked *"Do you think [*agent*'s name*] has good moral standards?" (1 = Completely; 7 = Not at all; r = .73). To measure anger and disgust towards the main agent, we asked participants to endorse on two separate scales how much the vignette made them feel like each of two photos of facial expressions, representing anger or disgust at 100% intensity (Beaupré , Cheung , & Hess, 2000), as well as scaled measures of the target emotions which asked participants to report three anger-related feelings (anger, outraged, furious) and three disgust-related (disgust, revolted, sickened) from 1 (Not at all) to 7 (Entirely) (Anger $\alpha = .91$; Disgust $\alpha = .95$). See Table 1 for descriptive

statistics. The measures of Experiment 1, and all subsequent experiments, can be found in supplementary material, document B.

Results

In the data structure, each participant x vignette combination was a separate case, containing the same measures. Each of the four dependent measures – wrongness, character, anger, disgust - was analyzed using two separate 2 x 2 mixed linear models, each of which crossed one of two moral code contrasts (harm vs purity; or harm vs pathogen) with a two-level context contrast (reality vs fiction). Participant was a random factor and moral code and context were fixed factors. This method accounted for the non-independence of each participant's responses to the within-participant vignettes.

Because generally similar patterns of results were identified across most of the moral judgment variables; we discuss the common pattern in terms of general "condemnation" while remarking on any individual variables that deviated from this pattern. See supplementary material, document C for the means and standard deviations of each experimental condition.

Harm vs Purity Contrasts

Consistent with our predictions, the contrast of fiction with reality yielded a series of significant interactions that indicated an evaluative asymmetry: a greater "fictive pass" that mitigated judgments of harm versus purity violations. Specifically, fictional harm, compared to real harm, showed reduced moral condemnation and anger, but this reduction was nonexistent or not as large when comparing fictional and real purity violations. These effects did not, however, apply to disgust, for which a non-significant Code x Context interaction ($p = .97$) indicated that

levels of disgust between real and fictional acts dropped by equal amounts across both code violations (see Figure 1).

Harm vs Pathogen Contrasts

We then tested whether pathogen violations showed asymmetry effects when compared to harm violations. As predicted, significant interactions across the reality-fiction contrasts indicated that pathogen acts, though less morally relevant than harm violations, were largely denied a pass. In other words, one who commits an act that evokes pathogen transmission in a fictional context is as condemnable as one who commits the acts in real-life. As with the harm-purity contrast, this asymmetry was not observed in disgust. It dropped equal amounts between real and fictional acts for both code violations (see Figure 2).

For moral judgments and anger, these results strongly suggest that acts that are disgusting because they threaten personal health, even when less morally relevant than harmful acts, are about as condemnable in fiction as in real-life. It should be noted that while harm violations were seen as more immoral than pathogen violation, the mean of the pathogen acts was closer to the scale midpoint of 4 ($M = 2.90$) than to the lowest scale point of 1. Effectively, some moral condemnation of these acts was present, so these effects cannot be easily explained by a floor effect.

Summary

Experiment 1 lent initial empirical support to the fictive pass asymmetry hypothesis. As expected, fictitious harm versus purity violations were granted more of a fictive pass (i.e. were less condemned), as indicated by a steeper slope from reality to fiction. This was true for general moral judgments, perceived moral character, and anger towards one who perpetrates harmful

behavior. By contrast, purity violations were denied a fictive pass to some extent, presenting a significantly less severe evaluative discrepancy between reality and fiction.

The one unexpected finding was in the exact nature of the lack of asymmetry for disgust. We confirmed that disgust would show a similar degree of fictive pass in both harm and purity violations. However, we also thought that the fictive pass effect for disgust would be low; but disgust actually showed an equal and significant drop for both types of violations, such that it was granted more of a fictive pass for purity violations than anger or overall judgment was. At this point, however, it is premature to conclude anything without the effect generalizing and replicating in further experiments.

Experiment 2

As an initial test of the fictive pass, Experiment 1, measured general morality and moral character, but there was not a clear distinction between act-based and character-based judgments (Uhlmann & Zhu, 2013). Experiment 2, therefore, used separate scales to distinguish between these types of judgments. We also dropped the pathogen condition in order to focus more centrally on highly morally relevant violations.

Furthermore, the within-participants design of Experiment 1 required participants to make joint evaluations of the different contexts in the same session. Because allowing explicit comparisons of the contexts may have distorted evaluations (Greenwald, 1976), Experiment 2 used an entirely between-subjects design in which each participant evaluated a single vignette.

In this second experiment, we used different vignettes to present harmful and impure acts. We also sought greater comparability between the fictional contexts, replacing the imagination condition with a video game condition, so that both were media products.

Method

Three hundred fifty-seven participants were recruited online (229 male; M_{age} = 31.5, SD_{age} = 9.7). Forty-four (12%) were excluded from analysis for reasons previously discussed.

Design

Experiment 2 employed a 2 (moral code: harm vs purity) x 2 (context of violation: reality vs fiction, collapsed from 3 conditions) between-subjects design. The questionnaire presented a single story setting that randomly presented one of the six possible combinations of conditions. Harmful acts violated another's rights, but were free of physical disgust (e.g. a young man deceiving an old woman for inheritance). Impure acts, by contrast, involved violations of bodily norms that were entirely consensual and free of harm (e.g. a young man having a consensual sexual relationship with an elderly woman).

Materials

The wording of the vignettes was manipulated so that the acts could be presented as occurring in a variety of contexts. For example, one of the purity vignettes in the context of real-life read, "Robert is a university student who owns a piercing gun. He goes to parties and enjoys giving genital piercings to anyone who wants one." In the context of being watched in a film the vignette read, "Robert watches a film about a university student who owns a piercing gun. The student goes to parties and enjoys giving genital piercings to anyone who wants one. Robert enjoys watching this film."

The emotion measures were unchanged from Experiment 1. Despite this, the mean scores of the anger items (α = .97) and the disgust items (α = .96) had an unusually strong correlation, r

= .87, $p < .001$. Because of this, we created a composite score of moral outrage. The act and character-based judgments were differentiated with one scale measuring the agent's moral character (Is [A*gent] "rotten inside"*?; Is [A*gent] immoral?; Is [Agent's soul impure?; Would* you say that [Agent] has good character?; Is [Agent] mainly a good or mainly a bad person?; α = .91) and another measuring the morality of the act he committed (*Is this a "rotten" thing to do;* Is this action morally blameworthy? Is this action deserving of punishment? Is this action immoral? α = .96). As with the emotion measures, however, the act and character scales were highly correlated ($r = .85$, $p < .001$) so a composite item of moral wrongness was created. The correlation of the moral wrongness and moral outrage composites was, however, also very high at $r = .83$, $p < .001$. As such, we created yet another composite that assessed negativity towards the described act.

The single resultant dependent measure – negativity -- was analyzed using an ANOVA that crossed the type of violation with the reality/fiction contrast, 2 (moral code: harm vs purity) x 2 (context of violation: reality vs fiction).

Results

For the measure of negativity, the main effects of code and context indicated that real acts and harmful acts were judged as relatively more negative that impure act and fictional acts. A significant Code x Context interaction also supported the fictive pass asymmetry hypothesis. Fictional harm was judged as less negative than real harm but fictional purity was judged about as negatively as its real-life counterpart (see Figure 3).

Experiment 3

Although the results of Experiment 2 did not offer as clear a distinction as hoped between moral judgments of act and character or between emotional responses, the findings once more lent support to the fictive pass asymmetry hypothesis, while generalizing to a different set of violations, contexts, and using a between-participants design with no explicit comparison between vignettes. It was found that evaluations of harm declined significantly from real to fictional contexts, but evaluations of purity violations stayed about the same across both contexts.

One limitation of Experiment 2 was that we used a different set of violations to generalize from Experiment 1, but without pretesting for equivalent moral wrongness. As it turned out, that set of purity violations was rated as overall less wrong than the harm violations. Mean condemnation of real and fictional purity violations alike was about halfway between the scale minimum of 1 and midpoint of 4, raising the possibility that a floor effect could be held responsible for the lesser difference between real and fictive contexts in purity. In fact, a similar if smaller interaction effect was found when the data were analyzed excluding all negativity mean responses less than 2, $F(1, 183) = 9.27$, $p = .003$, $\eta_p^2 = .05$, making it less plausible that a floor effect was completely responsible for our effects. However, the reduction in effect shows that it would be desirable to use moral situations with means closer to the midpoint. Therefore, Experiment 3 joined together all the fictive contexts tested so far, and used a new set of pretested moral situations that would all be near or above the scale midpoint in real-life condemnation.

Method

Pretest and vignette selection

The researchers wrote 42 vignettes that depicted violations of harm or purity. Participants (N = 132) recruited from Amazon's Mechanical Turk service were randomly presented with six vignettes. Each was followed by a forced choice facial expression agreement measure for anger and disgust using expressions from Beaupré et al., (2000). After this, self-report measures of anger, disgust, and moral wrongness were presented in a random order. These measures all ranged from 1 (not at all) to 7 (extremely). The average length of participation was 3.5 minutes and participants were compensated with $0.40.

Ultimately, we selected two harm vignettes and two purity vignettes that elicited the expected differential emotions (e.g. purity violations that were higher in disgust than anger) and that were similar in moral wrongness. There was no statistically significant difference in moral wrongness between the two harm vignettes, $t(26) = 0.96$, $p = .35$), the two purity vignettes, ($t(26) = 0.74$, $p = .47$), or among all four vignettes, $F(3,61) = 0.44$, $p = .73$.

Participants

321 participants (229 male; $M_{age} = 31.5$, $SD_{age} = 9.7$) were recruited from Mechanical Turk. Owing to similar measures and manipulations, anyone who had participated in Experiment 2 was not able to participate. Data of two participants (1%) were excluded for the previously stated reasons.

Design and Materials

Experiment 3 was analyzed using a 2 (moral code: harm vs purity) x 2 (context of violation: reality vs fiction, collapsed from 4 conditions) between-subjects design. Participants were randomly assigned to read and evaluate a single vignette that presented a harm violation or a purity violation occurring in one of the four different contexts (real-life, imagined, watched in a

film, performed in a video game). The harm vignettes described violations of autonomy (property destruction; verbal aggression) without any bodily moral norms violations. The purity vignettes described moral violations involving the body (sex with a dead chicken; bizarre bathroom behavior) that were free of harm to other persons. As before, the wording of these vignettes was manipulated to describe the acts as occurring in different contexts. For example, a harm code vignette in the context of real life read, "Sam shouted at his girlfriend because she did not have enough time to put on make-up before a date". In the context of played in a video game, the same vignette read, "Sam plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Sam can control his character to do in this virtual environment. In this video game, Sam controls a character that's the same age as he is. He controls his character to shout at his character's girlfriend because she did not have enough time to put on make-up before a date. Sam enjoys playing this video game". The dependent variables were unchanged from Experiment 2.

Our fictive pass asymmetry hypothesis remained the same. We predicted that harm would display a steeper drop in condemnation from real life to fiction but that purity violations would remain more stable across these contexts.

Results

The four anger items ($\alpha = .94$) and the four disgust items ($\alpha = .94$) were compiled into anger and disgust composite scores. The composites had a significant positive correlation with each other, but unlike in Experiment 2, the correlation was low enough ($r = .42$) that we could analyze each emotion separately.

As before, the act-based ($\alpha = .91$) and character-based ($\alpha = .94$) moral judgment items both had strong Cronbach's alphas and were turned into composite items. The correlation between these two scales was lower than in Experiment 2, but with a correlation of $r = .77$, these two items still shared 59% of their variance. Regardless, these items were analyzed individually so that we could begin to identify act- and character-based explanations for the fictive pass asymmetry hypothesis (see Table 3).

Main effects indicated that there was, in the main, little difference in character-based judgments between harm and purity violations. However, despite our pretesting real acts of harm ($M = 5.01$) were significantly more immoral than real acts of purity ($M = 4.01$), $F(1, 73) = 8.06$, $p = .006$. In line with pretesting, however, harm violations were associated most strongly with anger, and purity violations were associated most strongly with disgust. Furthermore, the baseline difference between real harm and purity acts did not plausibly mean that a floor effect on fictional purity acts could be held responsible for the fictive pass interaction. The mean for fictional purity, 3.49, was still much closer to the midpoint of 4 than to the scale minimum of 1.

Code x Context interactions for the reality-fiction contrasts supported a fictive pass asymmetry in act-based judgments. Fictional harm was more acceptable than real-harm but fictional purity was about as immoral as its real-life counterpart, even though real harm was condemned more than real purity violations were.

Unlike the results for our previous experiments, character inferences showed weaker fictive pass asymmetry effects, with no significant interaction. This was partly due to real harmful acts being seen as less indicative of bad character, relative to their moral condemnation.

Moral emotions, however, showed the expected main effects and, as in Experiment 1, partial evidence of a fictive pass asymmetry. Anger showed fictive pass effects by dropping significantly more for harm than for purity code violations. As in Experiment 1, however, disgust's reality to fiction drop was about the same, although not a significant decrease, for both code violations. In particular, the lack of a significant "pass" effect for disgust toward harm violations was more consistent with our initial predictions (see Figure 4).

Experiment 4[1]

Experiments 1 through 3 have shown evidence that fictional harm is, by and large, more acceptable than fictional purity and that when anger and disgust are distinct from one another (as in Experiments 1 and 3), anger shows a fictive pass asymmetry whereas disgust does not. Furthermore, judgments of character often were highly correlated with judgments of the acts. Although this is consistent with the explanation that fictive purity violations are disapproved of because they indicate bad character, it would be more theoretically useful to be able to contrast judgments of character against other facets of moral judgment that might not be expected to reflect the fictive pass asymmetry as strongly. Experiment 4 added a number of these facets.

Given that act and character in the foregoing experiment were very closely related, a more distinct aspect of moral judgment might be the desire to punish. In scenarios describing potentially harmful behavior, Cushman (2008) found that manipulating desire to harm produced the greatest effect on judgments of bad character; while manipulating the actual harmfulness of the consequences produced a more unique effect on recommendations to punish the perpetrator.

[1] Another experiment was conducted between Experiment 3 and Experiment 4. Because of its similarity to the present Experiment 4 we decided to place it in supplementary materials G as Experiment 4b.

Both desire and harmfulness were related to judgments of overall moral wrongness. In the context of our research, a fictional code violation might be seen as morally wrong and as revealing bad character, but less worthy of punishment than a real wrong due to its lack of consequence. We thus tested whether punishment, distinct from moral judgment, would respond more strongly to consequence than to signs of bad character.

A second innovation in this experiment explores the possibility that consuming immoral fiction is seen as having downstream consequences. Although we have speculated that engaging in impure fiction could be interpreted as a cue to an already bad character, another possibility is that consuming impure fiction is seen to actually cause bad character and bad behaviors. The Hays Code of 1930, which regulated the moral content of United States cinema, argued that films may "affect the moral standards of those who, through the screen, take in these ideas and ideals" or "inspire others with a desire for imitation" (Bynum, 2006), demonstrating concerns that media consumers might become corrupted by what they see, as well as emulating those behaviors. Of course, these two reasons are not mutually exclusive and might influence each other. We therefore included questions in Experiment 4 explicitly asking whether the acts described in the various conditions could worsen the consumer's character, and the extent to which the consumers may replicate this behavior. To avoid making these items seem tautological or non-sensual in the context of reality, the wording was slightly modified to fit each level of the context variable (e.g. from "*Do [thoughts/films/video games] like this corrupt one's character?*" to "Do these actions corrupt *one's character*?").

Method

Three hundred and fifty-two participants (195 male; M_{age} = 33.98; SD_{age} = 10.51) were recruited from Mechanical Turk. Four participants (1.1%) were excluded according to the previous stated criteria. Participants from the former experiments were not able to participate.

Design

The 2 (code violation: harm vs purity) x 2 (context of violation: reality vs fiction) between-subjects design was identical to that of Experiments 3 as were the vignettes that described the various immoral acts. The items that measured anger, disgust, and moral character were also unchanged. In an attempt to address the potential floor effects of the act-based judgments of the purity condition we clarified our measures of act condemnation. For each fictional contexts, the wording of the act-based judgments were slightly modified in order to clarify to participants that they were to be evaluating the fictitious act, not its real-life counterpart (e.g.: Is it morally blameworthy to [imagine this/do this in a video game/watch this in a film]?). We believe that this clarification will allow for a greater range of responses.

Materials

The new measures of consequence assessed the extent to which fictional acts caused one to become corrupt (a cause of bad character) as well as the likelihood that one would commit the acts in real-life. For example, participants rated items such as "Will [playing these sorts of video games/watching these sorts of films/imagining these sorts of things] make Sam a morally bad person? and "Will Sam do this in real-life because he [did it in a video game/watched it in a film/imagined doing it]?" (9 items; α =. 93). Separating these items into two subscales, one for consequences to character and one for behavioral consequences, revealed a correlation of r = .89. This gives further support for the assertion that character is useful as a cue to future behavior

(Tannenbaum et al., 2011,Study 2; Pizarro, et al., 2012), and supports our decision to analyze the two types of item as a single scale. Across the different contexts the wording of these items was set to match the fictional context that was being evaluated. Again, they were only presented to participants who read fictional acts.

Lastly, participants' desire to punish the offending agent was assessed with a single item that asked the extent to which the character in the vignette should be punished for his actions. This item was presented across all the contexts, real and fictional. Since the desire to punish is associated with the harmfulness of one's actions (Cushman, 2008), we expected this measure to make further distinctions between judgments of character and the consequences of engaging with fictional code violations.

All scales had strong reliability and the correlations, means, and alphas of each scale are listed in Table 4.

Results

The Fictive Pass Asymmetry

As with our former experiments, significant Code x Context interactions for act and character-based judgments showed that a fictive pass was given to harm code violations more so than purity code violations (see Figure 5). As expected, harmful behavior, compared to impure behavior, was relatively more acceptable in fiction than in real-life. This was also true of the desire to punish and for consequence ratings. However, because fictional acts were rated as equally consequential across both contexts, these effects are mostly due to real purity being seen as less consequential than real harm (Figure 6).

Since we did not change our manipulations, real acts of harm (M = 4.92) have once again been rated as more condemnable than real impurities (M = 4.26), $F(1, 84) = 11.52$, $p < .001$). However, our modified measures of act condemnation have allowed for a wider range of responses and we can now see that harm violations may start higher on the scale than purity violations, but they also end at a lower point (Figure 5). This steeper reality to fiction slope for harm, relative to purity, code violations disallows the possibility that our results have been explained by floor effects of the purity condition.

Moral emotions showed the expected main effect differences between anger and disgust in that purity, but not harm, code violations showed more disgust than anger. Only anger, however, showed a fictive pass asymmetry interaction. As in Experiment 3, the reality-to-fiction drop for disgust was about the same, and in fact non-significant, for both harm and purity.

Effects of mediated moderation

Our next goal was to test for mediation of the key Code x Context interaction on moral judgment (mediated moderation). The analysis was conducted with model 8 of the PROCESS macro (Hayes, 2012) at 10,000 iterations. Context (reality vs fiction) was the predictor variable, code (harm vs purity) was the moderator, moral wrongness was the outcome, and character judgments and future consequences were parallel mediators. See Figure 7 for a visualization of this model along with the unstandardized regression coefficients.

Indirect effects indicate that the effects of the Code x Context interaction on moral wrongness were significantly mediated by both character judgments (b = .44, SE = .21, 95% CI = [.10, .93]) and concerns of future consequences (b = .51, SE = .18, 95% CI = [.20, .89]). Moreover, the conditional direct effects indicate that judgments of purity code violations were

fully explained by the character-related mediators ($b = -.03$, $p = .89$), whereas judgments of harm violations were not entirely accounted for ($b = -1.28$, $p < .001$).

Thus, substantial variance in the complete fictive pass asymmetry effect can be explained by the fact that in fictional contexts, purity, as much as harm, code violations are seen as cue to a bad character as well as a cause of future corruption.

Experiment 5

Experiment 4 lent further support to the Fictive Pass Asymmetry hypothesis by once again showing an evaluative discrepancy between harm and purity code violations across real and fictional contexts. Fictional acts of harm were significantly less condemnable than their real life counterparts, but this gap was relatively smaller for impure acts. Furthermore, Experiment 4 explained these effects by indicating that purity, but not harm, code violations are seen as a cue to a bad character, as well as the cause of future corruption. This was true regardless of the fictional or real context that the act occurred in.

In Experiment 5 we pretested a new set of vignettes to increase the generalizability of our results and to address an issue with the pretesting of the vignettes used previously. Our previous pretesting had selected "purity" and "harm" scenarios based on their ability to elicit disgust more so than anger, and vice versa. However, this method of selection depended primarily on an outcome of relative emotional correspondences to these kinds of violation. Anger and disgust do not in and of themselves define the difference between scenario types, whether one interprets this as focused on the kind of moral principle violated (e.g. Graham et al ***), or the target of perceived harm (e.g. Schein Grey et al. ***; see below). We felt that a more valid method of pretesting would be to contrast acts that are seen as immoral and as harmful to other people,

versus immoral and not seen as harming other people. This minimal test would satisfy both the theoretical perspective that acts without harm to others are condemned because they violate a separate purity code of morality, and the perspective that they are condemned only because they are seen to harm entities beyond other people (such as nature, God or the self). Therefore, we asked how much each pretested vignette harmed other people and the extent to which it was morally wrong.

We also extended the measures of emotion in order to more fully distinguish between anger and disgust, both in the pretest and in Experiment 5 itself. These extended emotion measures consisted of metaphors of each emotion (e.g.: This makes my blood boil; This makes me feel like I will lose my appetite) and were to account for the possibility that participants are using disgust language to convey anger (Russell, Piazza, Giner-Sorolla, 2013). In conjunction with our facial expression agreement items, and word item measures of the target emotions and synonyms of each, these items were expected to further distinguish between anger and disgust empirically. See supplementary materials F for a full description of the pretest's method and results.

The perceived harmfulness of impurity

As mentioned in the Introduction, recent work has suggested that harm can be perceived in objectively harmless purity code violations (Gray, Schein, & Ward, 2014; Gutierrez & Giner-Sorolla, 2011) because victimless, yet immoral, acts conflict with one's template of dyadic morality that requires both an offender and a victim (Gray, Waytz, & Young, 2012). In the absence of a victim, Gray et al. (2014) argue that individuals will complete their dyadic template of wrongdoing by assuming that harm must have still occurred to an impersonal entity. Study 5

tested the possibility that the kind of acts we have heretofore characterized as "harm" can also be characterized as "harm to (specific) others" while "purity" acts can be characterized as harming other entities (e.g., harming the self (Chakroff, Dungan, & Young, 2013) or harming nature (Gutierrez & Giner-Sorolla, 2011). To reinforce the uniqueness of our harm and purity vignettes, we have introduced new items to measure the perceived harmfulness of each code violation to a variety of targets. We predict that participants' imputations of harm will significantly differ between our manipulations and show a clear distinction between harmful and impure behaviors.

A heterogeneous assessment of moral character

In our preceding experiments, our measures of moral character using simple items such as, "Is Sam mainly a bad person" may have overlooked the heterogeneous nature of moral character, which goes beyond mere evaluation of a person as good or bad to encompass specific positive prosocial traits (Goodwin, Piazza, & Rozin, 2014). To address this, we have included new items such as warmth, fairness, empathy, integrity, and abnormality (Goodwin et al., 2014; Uhlmann, Pizarro, & Diermeier, 2015) in order to more thoroughly encapsulate the scope of moral character. These items also have the advantage of being more independent of the purity construct, compared to our previous items such as "*Is Sam's soul* impure" or "is Sam rotten inside".

Desire as an alternative explanation of the asymmetry

The final innovation of Experiment 5 sought to identify explanations of the fictive pass asymmetry effects beyond those of character from the previous experiment. Russell and Piazza (2015, Study 4) found that bizarre sexual desires were condemned as much as when the desired act was actually committed. These findings suggest that asymmetries in perceived desire might

drive the fictive pass asymmetry; perhaps people who consume fictional impurity will be seen to actually desire to perform the act more so than people who consume fictional harm. We tested this idea by measuring perceived desire to commit the act in the target of the scenario, and testing desire as a mediator of the fictive pass asymmetry, as we did with character factors in the previous study.

Method

Four hundred and eighty four United States residents (284 male; $M_{age} = 34.77$; $SD_{age} = 10.36$) were recruited from Amazon's Mechanical Turk service. The questionnaire included an attention checking question. If participants answered it incorrectly then the survey automatically directed them to the debrief page and recruited a new participant in their place.

Design, Materials, and Procedures

As with our former experiments, Experiment 5 was a 2 (code violation: harm vs purity) x 2 (context of violation: reality vs fiction) between-subjects design. The three fictional contexts (imagined, watched in a film, performed in a video game) were collapsed into to form a single level of fiction. As in the previous experiments the vignettes' text was manipulated to present the acts as occurring in different contexts. For example, one of the purity code violations in the context of real-life read, "Sam has sex with a frozen chicken before cooking it and eating it for dinner. Sam enjoys doing this". In the context of imagination, however, the vignette read, "Sam imagines that he has sex with a frozen chicken before cooking it and eating it for dinner. He enjoys imagining this". All dependent items were measured on a Likert-type scale from 1 (not at all) to 7 (entirely).

Emotion Measures

The emotion word-item measures and the facial expression agreement items were unchanged from our previous experiments and as previously mentioned, further distinctions between anger and disgust were made with three metaphors for each target emotion (i.e.: this makes my blood boil; this makes me lose my appetite). All items were collapsed into two composite variables. The anger items ($\alpha = .95$) and the disgust items ($\alpha = .96$) were reliable, had a relatively low correlation ($r = 0.50$) and thus, were analysed separately.

Act Judgments

Act-based judgments (3 items; $\alpha = .91$) were unchanged from the previous experiment. As before, the wording of these items was modified depending on what context the act was presented as occurring in (e.g.: Is this morally blameworthy?/Is it morally blameworthy to [imagine this/watch this in a film/perform this in a video game]?).

Character Judgments

Drawing from the literature on moral character (e.g. Goodwin et al., 2014; Uhlmann, Pizarro, & Diermeier, 2015), thirteen items measured different facets of moral character. Participants reported the extent to which they perceive the main agent as abnormal, twisted, perverse, deviant, trustworthy, fair, loyal, empathetic, reliable, warm, and having integrity. An exploratory factor analysis with a maximum likelihood extraction and a promax rotation indicated that these items loaded onto two distinct factors. One factor contained items that related to positive and praiseworthy character traits (i.e.: warmth, loyalty, empathy, fairness). Items that loaded on the second factor related to negative and abnormal traits (i.e.: perverseness, deviance, indecency). The items of these factors had strong reliabilities (both α's = .94) and shared a correlation of $r = .29$, $p < .001$, thus allowing us to collapse them into two variables:

moral character and abnormal character. All items were coded so that higher numbers reflected more immorality or abnormality.

Measures of Harm

Perceptions of harm (Gray et al., 2012; Gray et al., 2014; Gutierrez & Giner-Sorolla, 2011) was measured towards three different entities. The items below are displayed in the context of reality but these items were modified to fit each fictional context.

Social harm: Five items measured the perceived harm the agent's actions caused to other individuals and to the community at large (*Do you think that Sam's actions caused* [psychological/physical/emotional] harm to anyone other than himself? *Do you think Sam's* actions violated the rights of anyone other than himself? Do *you think that Sam's actions caused* harm to society at large? $\alpha = .90$.)

Self harm: Three items measured the perceived self-harm of the agent's actions (e.g.: Do *you think that Sam's actions caused* [psychological/physical/emotional] harm to himself? $\alpha = .80$).

Natural Harm: Two items ($r = .73$) measured the perceived harm the agent's actions caused to the natural order (e.g.: *Do you think that Sam's behaviour caused damage to the natural order of things? Did Sam's actions violate any laws of nature?*)

The items of these three scales were subjected to an exploratory factor analysis with a maximum likelihood extraction and a promax rotation. It was indicated that these items loaded onto two distinct factors. The natural harm and self harm items (5 items: $\alpha = .88$) loaded on one factor, and were averaged into the variable non-social harm. The social harm items formed one

factor (5 items; $\alpha = .90$). This loading empirically supports the idea that acts seen to harm other people, whether in the individual or aggregate, form a different class than acts not seen to harm other people, whether this harm is defined as to the self or to nature.

Desires

Three items ($\alpha = .96$) measured the main agent's perceived desire to commit the described act (e.g.: Do you think that Sam [did this/imagined this/watched this in a film/controlled his character to do this in a video game] because he desires to actually get into a fight with another man and punch him in the face?). Similar to previous items, the wording was modified to fit the different levels of the context variable.

Results

The Fictive Pass Asymmetry

As in our former studies, significant Code x Context interactions revealed fictive pass asymmetry effects for act judgments and (reversed) judgments of positive moral character. There was a significantly greater reality to fiction drop in moral wrongness for harm code violations than for purity code violations (Figure 8). Despite their equality in pretesting, real acts of harm were significantly more immoral than real impurities, $F(1, 122) = 17.75$. $p < .001$, $\eta_p^2 = .13$. In spite of this, the moral wrongness of harm violations both started higher and ended lower on the scale (Figure 8), while the wrongness of fictitious purity violations was close to the scale midpoint. Both these features argue against the possibility that the difference in baseline morality between violation types presents a problem in interpreting the fictive pass asymmetry due to floor effects on fictitious purity violations. If anything, the floor effect is on fictitious harm violations (with a mean near 2) and would work against, not for the fictive pass asymmetry.

Main effects of abnormal character indicated that real acts compared to fictional acts, $F(1, 483) = 83.74$, $p < .001$, $\eta_p^2 = .15$, and impure acts compared to harmful acts, $F(1, 483) = 102.81$, $p < .001$, $\eta_p^2 = .18$, were most indicative of character abnormality. The Code x Context interaction ($p = .11$) was not statistically significant but close enough to marginal significance that it should not be discounted from future analyses and experiments.

As in Experiment 4, main effects of anger and disgust showed that harm, but not purity, code violations showed more anger than disgust and vice versa but only anger showed a fictive pass asymmetry effect. Consistent with Experiments 3 and 4, the reality-to-fiction drop for disgust was non-significant for both code violations (Figure 8).

Main effects of desire indicated that purity code violations were more indicative of desire than harm violations, $F(1, 453) = 4.82$, $p = .03$, $\eta_p^2 = .01$ and that real acts, relative to fictional acts indicated more true desires, $F(1, 453) = 147.29$, $p < .001$, $\eta_p^2 = .25$. However, desire did not show a fictive pass asymmetry interaction ($p = .57$).

The harm of "harmless" (to others) impurities

New measures of harmfulness addressed the possibility that our harm-purity distinction could alternatively be described as a distinction between acts that harm another person, and acts that harm some other entity (Gray et al., 2014).

The effect of our manipulations on the perceived type of harm was shown in a significant Harm Type x Code Violation x Context interaction, $F(1, 450) = 107.75$, $p < .001$, $\eta_p^2 = .19$. More specifically, a Harm Type x Code Violation interaction, $F(1, 450) = 321.01$, $p < .001$, $\eta_p^2 = .42$ indicated that harmful acts evoked stronger perception of social harms ($M = 3.34$) than non-

Comment [J1]: Added 8/6/16

social harms ($M = 2.26$, $p < .001$), and impure acts more strongly evoked non-social harms ($M = 3.28$) than social harms ($M = 2.19$, $p < .001$). This shows that our "harm/purity" distinction can also be characterized in terms of harm to social versus non-social entities.

Like the categories of harm and purity, different kinds of harm also were associated with different moral emotions. When controlling for disgust, social harm ($b = .37$, $p < .001$) more so than non-social harm ($b = .05$, $p = .32$) predicted anger; when controlling for anger, non-social harm strongly positively predicted disgust ($b = .71$, $p < .001$), unlike social harm ($b = -0.30$, $p < .001$).

Effects of mediated moderation

An analysis of mediated moderation was conducted with the PROCESS macro's 8th model (Hayes, 2012) at 10,000 iterations. Context (reality vs fiction) was the predictor variable, code (harm vs purity) was the moderator, moral wrongness was the outcome, and judgments of character morality, character abnormality, and desires, were all set as parallel mediators. See Figure 10 for a visualization of this model as well as the unstandardized regression coefficients.

The mediator's indirect effects of the key Code x Context interaction on moral wrongness were significantly mediated by judgments of moral character ($b = 0.18$, $SE = .05$, $CI = [0.10, 0.30]$) and judgments of abnormal character ($b = 0.14$, $SE = .05$, $CI = [0.06, 0.25]$), but not desire ($b = -0.06$, $SE = .04$, $CI = [-0.15, 0.02]$). Furthermore, the conditional direct effects indicated that the evaluations of purity code violations were fully explained by the character-related mediators ($b = -0.05$, $p = .60$) whereas social harm violations were not ($b = -0.50$, $p < .001$).

Expanding on the findings of our former experiments, the present experiment has provided a more thorough picture of the fictive pass asymmetry effects. Variance in the Code x

Context interaction on moral wrongness is explained by the fact that fictional purity, more so than fictional harm, code violation signals one as an abnormal person and an immoral person. The fictive pass asymmetry effects on moral wrongness are not, however, explained by the presumption of desires. To engage with fictional impurities does not imply any desire to commit the act in real life any more so than engaging with fictional harm to others does.

Meta-analyses of the experiments

In reporting analyses of all experiments we have collapsed the various fictional contexts (imagination, watched in a film, performed in a video game) into a single level, because we did not have any specific hypothesis regarding the fictive pass effects between these contexts. Table 6 in supplementary material, document D shows the specific Code x Individual fictive-context interactions across all five experiments. While this table can satisfy curiosity regarding any specific interactions, it is hard to draw any overarching conclusions from such a large display.

To condense and systematically analyze these results, we conducted a first meta-analysis of experiments 3, 4, and 5 in order to get a more holistic understanding of the fictive pass effects by context across our experiments. This meta-analysis examined whether any given context may be most or least responsible for the effects of the fictive pass asymmetry. Experiments 1 and 2 were not included in this analysis because they did not contain the full set of fictive contexts that ended up being included in Experiment 3 through 5.

The meta-analysis was conducted using downloadable meta-analysis macros for SPSS (Wilson, 2005). To prepare the data we obtained the mean difference in moral wrongness of each reality vs fictive context contrast for both harm and purity code violations from Experiment 3, 4, and 5. The results (Figure 11) indicate that the moral wrongness difference between real and

fictional contexts is greater for harm that it is for purity. This reinforces our consistent findings that harm, more so than purity, has greater influence on the effects of the fictive pass asymmetry.

As a rule of thumb, confidence intervals that do not overlap by more than 25% are considered to be significantly different from one another (Cummings & Finch, 2005). In general, the highest reality-fiction differences were found among video games, and the lowest among imagination. But all three contexts meta-analytically showed the critical asymmetry, in that the reality-fiction difference for harm was different from that for purity. Relatively speaking, the strongest asymmetry was seen for film, but overall, the fictive pass asymmetry was reliably found in all three fictive contexts studied.

We conducted a second meta-analysis of Experiments 3, 4, and 5 that focused on a different question: the overall effects of the fictive pass asymmetry on anger and disgust, collapsing as before the fictional contexts into a single level. To prepare the data, we calculated the reality vs. fiction mean difference for both anger and disgust across harm and purity code violations. Experiments 1 and 2 were excluded from analyses because Experiment 1 had a within-participants design that was hard to compare with the others, and Experiment 2 did not show sufficient differentiation between anger and disgust.

The results of this meta-analysis (Figure 12) illustrate that the overall effects are in line with our expectations and show the fictive pass asymmetry effects being stronger for anger than for disgust. Specifically, harm and purity scenarios showed about the same amount of decline in disgust from reality to fiction, but harm scenarios showed much more decline in anger than purity scenarios did. In comparing confidence intervals, the fictional mitigation of anger at harm was greater than the other three effects, which did not differ from each other.

Discussion

The fictive pass asymmetry

The results of these five experiments have supported our fictive pass asymmetry hypothesis by demonstrating that fictional contexts mitigate moral evaluations of acts that harm other people, more so than "purity" violations that are seen as harming only the self or abstract entities. Experiment 1 provided initial support by demonstrating that one who engages with fictional acts that harm others is seen as less immoral, less bad of a person, and evokes less anger than one who acts harmfully in real-life; while for purity code violations, the evaluative discrepancy between reality and fiction was relatively less extreme. Experiments 2 and 3 found similar effects and gave additional support to the fictive pass asymmetry hypothesis via methodological improvements and by expanding upon the fictional contexts that the code violations occurred in. Experiments 4 and 5 distinguished between two roles of fictive activity: as a cue to bad character and as a perceived cause of bad character and actions. These also found the fictive pass effect, and further showed that while both roles contributed to the asymmetry effect, the cue role was stronger, or about equal to, the cause role; in other words, people gave consumers of harmful fiction a "pass" because, unlike impure fiction, it was not seen as indicating anything bad about their moral character.

Although our hypothesis was phrased in terms of a difference between differences, it should also be noted that in general, this interaction took a specific form -- harm scenarios in real life were usually rated as more severe than purity scenarios, while in fiction purity tended to be rated as equivalent or worse than harm. This occurred even though we tried our best to pretest harm and purity scenarios that would be seen as equally wrong in real life, which may point to

the simple fact that in our participants' cultural context, harm violations are more condemnable than purity violations overall. This effect coexisting with the interaction and produced this specific pattern of means, which is still compatible with the idea that fiction leads to a stronger reduction in condemnation for harm versus purity. However, it is true that most of our studies did not literally find that purity in fiction would be condemned more than harm.

In those experiments for which anger and disgust were distinguishable from one another, anger categorically demonstrated fictive pass effects. Disgust, although less consistently, most often showed equal effects (or non-effects) between harm and purity code violations; and this difference between emotions was confirmed by the meta-analysis. In other words, for harm violations, disgust behaved differently than anger, showing less of a drop in fictive contexts; so that, when targets fictionally harmed someone, the prevalent reaction towards them tended to be disgust rather than anger. It may be, then, that disgust at fictional harm serves the purpose of evaluating the actor's character, even if there is no actual bad behavior or harmful action to be angry at. The role of disgust as a mark of character even in the absence of condemnable actions has been remarked upon (e.g. Miller, 1997; Rozin, Haidt, McCauley, 2008), but awaits further empirical confirmation.

Theoretical implications

The results of these experiments have demonstrated how the contexts that surround specific norm-violating acts influence how we morally evaluate these acts and the individuals involved. It is perhaps not surprising that fictional contexts should mitigate judgment of any immoral act, if one takes a purely utilitarian and consequentialist position: that right and wrong inhere only in the outcomes of the act. What is more noteworthy is that this mitigation is reduced

for violations of purity moral codes, supporting existing evidence that such codes are more related to judgments of moral character (Chakroff & Young, 2014). Furthermore, our findings indicate that beliefs about future behavior are intertwined with beliefs about effects on character. This suggests that character morality is somewhat rooted in long-term utilitarian concerns. Moreover, these experiments support and expand upon the rigid nature of purity code evaluations, compared to the relative flexibility of harm code evaluations. Former research has demonstrated how these effects occur in real-life scenarios (Gutierrez & Giner-Sorolla, 2007; Rozin, Millman, & Nemeroff, 1986), but our research has demonstrated how these principles hold true even in fictional contexts.

These claims, however, are hindered by the fact that our sample consisted of people living in the Western English-speaking world. As such, we cannot fully address the extent to which these effects would apply across different cultures[2]. While early work in cross-cultural morality did put forth the idea that violations of harm are universally immoral (Turiel, Killen, & Helwig, 1987). Haidt, Koller, & Dias (1993) contested the exclusive wrongness of harm by suggesting that, "The domain of morality appears to vary cross-culturally" (pg. 625) and more recently, work has demonstrated how culture is a critical facet of morality (Graham, Meindl, Beall, Johnson, & Zhang, (2015); Guerra & Giner-Sorolla, 2010; Vauclair & Fischer, 2011) and more specifically, the moralization of entertaining thoughts of immoral behavior can substantially vary between cultures (Cohen & Rozin, 2001). The amount of variability that is

[2] In fact, differences are visible between film ratings systems in the United States and Scandinavia; the Scandinavian system places a relatively greater weight on controlling the portrayal of violence versus sex (Price, Palsson, & Gentile, 2014). While sex and violence in films are not perfectly translatable to purity and harm violations, this demonstrates how different cultures can differently evaluate these concepts.

introduced by cross-cultural differences poses challenges while trying to ascertain universal truths about moral judgment. In spite of this, we believe that the fictive pass asymmetry does lend strong empirical support to the casually observed discrepancy between the appropriateness of fictional harm and purity codes, a cross-cultural study would be needed to assess the true generalizability of this work.

Similarly, one must consider the ecological validity of our results, and the extent to which our vignettes are truly representative of the types of acts that are commonly portrayed in fiction. In these experiments, the scenarios presented a experimental control, but perhaps at the cost of ecological validity. This is because media products rarely display acts that neatly violate a single moral code. For instance, harm violations usually manifest themselves in the form of violence. This violence, however, may infringe on the purity domain by presenting blood or gore. By contrast, purity code violations as they have been portrayed in these scenarios are rarely depicted in popular media. Indeed, examples such as the controversial modification to Grand Theft Auto demonstrate that sexual content does not need to be particularly abnormal in order to be controversial as an element of fiction. One way to explain these examples as purity violations is that it is the public and available depiction of an activity seen as sacred, such as sex, that leads to moral opprobrium, even if the act itself is not seen as immoral in the appropriate context. Another possibility is that the thought of fictional depictions of even acceptable sexual activities in the hands of children, through such media as books, games, films or comics, brings up concerns for their purity and innocence.

In fact, research on acceptable “community standards” of fiction has found that sexually explicit content intended for adults is seen as permissible, so long as minors are not involved and there are no depictions of sexual violence or fetishism such as bondage (Linz et al., 1995). While

this is problematic for some of the examples that we have used to contextualize this research, it more importantly supports our findings by demonstrating that the most condemnable acts are ones that involve the body in bizarre and unnatural ways. Most of all, however, it highlights the need of future research that would explore the fictive pass using depictions of acts that are plausibly encountered in real life and across various actual fictional contexts.

We believe that our research also sheds light on applications of moral psychology to media regulation. Organizations such as the USA's MPAA, North America's ESRB or Europe's PEGI are responsible for giving standardized ratings of age appropriateness to media products so that consumers and parents can be aware of their content. Interestingly enough, the general framework of these organizations' published criteria falls in line with the effects of the fictive pass asymmetry (esrb.org; pegi.info). Fictional acts of harm (mostly violence in the case of media products) are deemed appropriate for much younger ages than content that may be considered impure (such as sexual content, or other morally impure behavior such as drug use and gambling). Graphic depictions of blood and bodily destruction also merit older age ratings, again possibly due to the disgust and purity concerns that such displays bring up. These organizations do not offer any scientific explanations for their, perhaps intuitive, decision making. These experiments can therefore explain and justify their criteria as reflecting public opinions about the acceptability of fictional acts, both in our findings and in the possible extension of our methods to parent and community samples.

In closing, this current set of experiments has shown us that one may be given a pass for enjoying violent video games and films, or having aggressive thoughts towards another individual. Consequently, these fantasies may be seen as relatively benign and nonconsequential. When the fictional acts, however, involve a bizarre and socially unacceptable use of the body,

then they are not granted the same pass that is given to fictional social harm. Not only do these acts signal a poor character, they are seen as a cause of future indiscretions. As it turns out, not all fiction is treated equally and, while it is all make-believe, impure fiction is associated with very real consequences.

References

Beaupré , M. G. , Cheung , N. , & Hess , U. (2000) The Montreal set of facial displays of emotion [Slides] Available from Ursula Hess, Department of Psychology, University of Quebec at Montreal, Montreal, Quebec, Canada.

Brown v. Entertainment Merchants Association. 564 U.S. (2011)

Bynum, M. (2006) The motion picture production code of 1930 (Hays code). Retrieved from http://www.artsreformation.com/a001/hays-code.html

Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. Journal of personality and social psychology, 81(4), 697-710. http://dx.doi.org/10.1037/0022-3514.81.4.697

Chakroff, A., Dungan, J., & Young, L. (2013). Harming ourselves and defiling others: What determines a moral domain? PloS One, 8(9): e74434. doi: 10.1371/journal.pone.0074434

Chakroff, A. & Young, L. (2015) Harmful situations, impure people: An attribution asymmetry across moral domains. Cognition, 136, 30-37. doi: 10.1016/j.cognition.2014.11.034

Cumming, G., & Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. American Psychologist, 60(2), 170. doi: 10.1037/0003-066X.60.2.170

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. Cognition, 108(2), 353-380. doi:10.1016/j.cognition.2008.03.006

Giner-Sorolla, R., Bosson, J. K., Caswell, T. A., & Hettinger, V. E. (2012). Emotions in sexual morality: Testing the separate elicitors of anger and disgust. Cognition & Emotion, 26(7), 1208-1222. doi: 10.1080/02699931.2011.645278

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. Journal of personality and social psychology, 106(1), 148-168 . http://psycnet.apa.org/doi/10.1037/a0034726

Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. Current Opinion in Psychology, 8, 125-130. doi: 10.1016/j.copsyc.2015.09.007

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. doi: 10.1037/a0021847

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. Journal of Experimental Psychology: General, 143(4), 1600-1615. doi: http://psycnet.apa.org/doi/10.1037/a0036149

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. Psychological Inquiry, 23(2), 206-215. doi: 10.1080/1047840X.2012.686247

Gray, K., & Wegner, D. M. (2010). Blaming God for our pain: Human suffering and the divine mind. Personality and Social Psychology Review, 14(1), 7-16. doi: 10.1177/1088868309350299

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? Psychological Bulletin, 83(2), 314. doi:10.1037/0033-2909.83.2.314

Guerra, V. M., & Giner-Sorolla, R. (2010). The community, autonomy, and divinity scale (CADS): A new tool for the cross-cultural study of morality. Journal of Cross-Cultural Psychology, 41(1), 35-50. doi: 10.1177/0022022109348919

Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. Emotion, 7(4), 853-868. doi: 10.1037/1528-3542.7.4.853

Gutierrez, R., & Giner-Sorolla, R. (2011). Disgusting but harmless moral violations are perceived as harmful due to the negative emotions they elicit. Revista de Psicologia Social, 26(1), 141-148. doi: 10.1174/021347411794078381

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? Journal of Personality and Social Psychology, 65(4), 613-628. doi:10.1037/0022-3514.65.4.613

Hayes, A. F. (2012) SPSS PROCESS documentation. Retrieved from http://www.processmacro.org/download.html

Iwan, C. (2014) *Llanbradach headteacher's warning after pupils as young as six act out* drug and rape scenes from Grand Theft Auto. Retrieved from http://www.southwalesargus.co.uk/news/11001937.Warning_after_Valleys_pupils_as_young_as_six_act_out_rape_scenes_from_Grand_Theft_Auto

Leone, R. (2004). Rated sex: An analysis of the MPAA's use of the R and NC-17 ratings. Communication Research Reports, 21(1), 68–74. http://doi.org/10.1080/08824090409359968

Linz, D., Donnerstein, E., Shafer, B. J., Land, K. C., McCall, P. L., & Graesser, A. C. (1995). Discrepancies between the legal code and community standards for sex and violence: an

empirical challenge to traditional assumptions in obscenity law. Law and Society Review, 127-168. doi: 10.2307/3054056

Miller, W. (1997). The anatomy of disgust. Cambridge, Mass: Harvard University Press.

Olson, R. (2014) A look at sex, drugs, violence, and cursin in film over time through MPAA ratings. Retrieved from http://www.randalolson.com/2014/01/12/a-look-at-sex-drugs-violence-and-cursing-in-film-over-time-through-mpaa-ratings

Pizarro, D. A., Tannenbaum, D., & Uhlmann, E. (2012). Mindless, harmless, and blameworthy. Psychological Inquiry, 23(2), 185-188. doi:10.1080/1047840X.2012.670100

Price, J., Palsson, C., & Gentile, D. (2014). What matters in movie ratings? Cross-country differences in how content influences mature movie ratings. Journal of Children and Media, 8(3), 240-252. doi: 10.1080/17482798.2014.880359

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. Journal of Personality and Social Psychology, 50(4), 703-712. doi:10.1037/0022-3514.50.4.703

Rozin, P., Haidt, J., & McCauley, C. R. (2008) Disgust. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett. (Eds.), The Handbook of Emotions (3rd ed). (pp. 757-776). New York, NY: Guilford Press

Russell, P. S., & Giner-Sorolla, R. (2011a). Social justifications for moral emotions: When reasons for disgust are less elaborated than for anger. Emotion, 11, 637–646. doi:10.1037/a0022600

Russell, P. S., & Giner-Sorolla, R. (2011b). Moral anger is more flexible than moral disgust. Social Psychological and Personality Science, 2, 360–364. doi:10.1177/1948550610391678

Russell, P. S., Piazza, J., & Giner-Sorolla, R. (2013). CAD revisited effects of the word moral on the moral relevance of disgust (and other emotions). Social Psychological and Personality Science, 4(1), 62-68. doi: 10.1177/194855061244291

Russell, P. S., & Giner-Sorolla, R. (2013). Bodily moral disgust: What it is, how it is different from anger, and why it is an unreasoned emotion. Psychological Bulletin, 139(2), 328-351. doi:10.1037/a0029319

Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. Journal of Experimental Social Psychology, 47(6), 1249-1254. doi:10.1016/j.jesp.2011.05.010

Thompson, K. M., & Yokota, F. (2004). Violence, Sex, and Profanity in Films: Correlation of Movie Ratings With Content. Medscape General Medicine, 6(3), 3

Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, functions, and vagaries. In J. Kagan & S. Lamb (Eds.), The emergence of morality in young children (pp. 155-243)**.** Chicago: University of Chicago Press.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. Perspectives on Psychological Science, 10(1), 72-81. doi: 10.1177/1745691614556679

Uhlmann, E. L., Zhu, L., & Diermeier, D. (2014). When actions speak volumes: The role of inferences about moral character in outrage over racial bigotry. European Journal of Social Psychology, 44(1), 23-29. doi: 10.1002/ejsp.1987

Vauclair, C. M., & Fischer, R. (2011). Do cultural values predict individuals' moral attitudes? A cross-cultural multilevel approach. European Journal of Social Psychology, 41(5), 645-657. doi: 10.1002/ejsp.794

Wilson, D. B. (2005). Meta-analysis macros for SAS, SPSS, and Stata. Retrived November 6, 20015 from http://mason.gmu.edu/~dwilsonb/ma.html

Young, G., & Whitty, M. T. (2011). Should gamespace be a taboo-free zone? Moral and psychological implications for single-player video games. Theory & Psychology, 21(6), 802-820. doi: 10.1177/0959354310378926