THE FICTIVE PASS ASYMMETRY: CONDEMNATION OF HARM, BUT NOT PURITY, IS MITIGATED BY FICTIONAL CONTEXTS

Thesis submitted for the degree of

Doctor of philosophy

At

University of Kent

By

John Sabo

School of Psychology

University of Kent

September, 2016

Word Count: 38,475

ACKNOWLEDGEMENTS

I would like to extend my thanks to the many people whose patience, effort, and compassion has help me throughout the journey that has been my doctoral studies.

Special thanks is owed to my supervisor, Roger Giner-Sorolla. Roger has been by my side from the very start of my MSc degree to the end of my PhD. Without his patience, advice, and guidance my studies would have been much more turbulent than they were. I do not think that every PhD supervisor would tolerate unannounced drop-ins or helping a frantic student with meeting last minute deadlines, but Roger never let me down. My success is largely owed to his dedication and I could not be more grateful.

Additionally, there are a number of others who helped me along the way. Darren, has been a constant companion since day one. It has been a pleasure sharing an office, bouncing ideas off of each other, helping each other with research, and blowing off steam together over the weekends. My parents and my sister have given me unwavering support throughout my studies and while I do not call as often as I should, they need to know that I would never have gotten this far without them. Lastly, Nina, more so than anyone else, has seen firsthand the toll that work can have on me and her support has been beyond belief. I shudder to think what this experience would have been like without her by my side.

To all my family and friends, thank you! This work is dedicated to you.

ABSTRACT

Is there a double standard when it comes to the moral acceptability of fiction that encourages the imagination of acts that violate moral norms of harm and moral norms of purity? Observations of ethics, legal proceedings, and public reactions to different types of media seems to suggest so. Over six experiments this phenomenon, coined the fictive pass asymmetry, will be tested. The fictive pass asymmetry hypothesis proposes that fictional contexts including imagination, film, and virtual environments, will mitigate the condemnation of harm code violations more so than purity code violations. In other words, fictional representations of harm are given a "fictive pass" in moral condemnation, but the fictional representation of purity code violations that involve an abnormal use of one's body are denied a pass, and thus evaluated more similarly across real and fictional contexts.

Chapters 1 through 3 introduce the fictive pass asymmetry and review the literature that provide its theoretical framework. Chapter 4 presents three experiments that establish initial evidence in support of the fictive pass asymmetry effects. Experiment 1 presented participants (N = 431) with vignettes that described agents committing either sexual acts or violent acts that were described as occurring in real life, being performed in a video game, or watched in a film. Experiments 2 and 3 (N = 360 and N = 321, respectively) systematically improved methodology by expanding upon the fictive contexts and creating manipulations based more strictly on the moral psychology literature. Chapter 5 presents experiment 4 (N = 312) and experiment 5 (N = 352) which deepened the understanding of the fictive pass asymmetry effects by using mediation analyses to demonstrate how the

perceived wrongness of fictional purity code violations can be explained by the extent to which they signal poor moral character. Lastly, chapter 6 contains a final experiment (N = 484) and a series of meta-analyses. The final experiment considers fictive pass asymmetry effects in relation to an opposing theoretical framework, validates a number of manipulations, and tests the presumption of desire as an alternate explanation of fictive pass asymmetry effects. Finally, the meta-analyses aggregate the data of these experiments to highlight the robustness of the fictive pass asymmetry effects. Chapter 7, the concluding chapter, reviews the experiments and discusses the results in regards to theories of anger and disgust, moral theories of act and character, as well as the fictive pass asymmetry's implications in media use and regulation.

Contents

CHAPTER 1

GENERAL INTRODUCTION

"*Until I went online and checked the content of the 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)

Can behaviors that are committed within a fictional context be perceived as immoral? If so, is there a double standard regarding the permissibility of fictional representation sex versus violence? The above quote was spoken by a primary school teacher in regards to the sexual content of the 2013 video game *Grand Theft Auto*; the most recent installment in a line of controversial video games in which players find themselves committing a number of illicit acts ranging from police murder to gang killings and drug dealing. The teacher's words suggest that he has a rather blasé attitude towards the video game's violent content, even to the extent that he has normalized pupils' engagement with fiction that encourages them to imagine acts of violence. However, when he realized that the game also allowed its players to engage with fictional sexual content, red flags were raised and his disapproval of the game became more vocal.

This quote highlights a larger point that will be explored within this thesis: not all fictional behaviors are evaluated equally. From a purely consequentialist point of view, it is puzzling that this schoolteacher evaluates the appropriateness of fictional violence and sex to different degrees. After

all, acts that are performed in video games are all equal in the sense that they are make believe and do not result in any actual harm. So then why is it that this school teacher has such a cavalier attitude towards the fictional violence, framing it as entirely normal and appropriate, while being so condemnable of the sexual content?

Unsurprisingly, the acceptance of violent video games is widespread and it goes well beyond this one school teacher. The general popularity of violent video games is demonstrated by their sheer abundance and massive popularity. Generally speaking, a survey of the video gaming habits of American adolescents suggested that 97% of American young adults play video games, 31% of them play games on a daily basis, and 50% of the sample played games that had violent content (Lenhart, et al.,2008). Video game play is similarly prevalent amongst adults (Lenhart, Jones, & Macgill, 2008) and their popularity is further highlighted via their financial success in that they typically outsell their non-violent counterparts, and that video gamers rate them as being more enjoyable to play (Greitemeyer & Mügge, 2014).

Violent video games are both popular and in high demand, but all of this is not to say that they are accepted with open arms. Much empirical work has demonstrated how violent video game play is associated with an increase in aggressive cognitions, attitudes, and behaviors (Bensley & Van Enwyk, 2001; Greitemeyer & Mügge, 2014). For instance, the previously mentioned *Grand Theft Auto* series are often times condemned for allowing people to virtually engage in a wide variety of criminal activities that range from car theft, to dealing drugs, and to killing innocent people on the street.

Some games even take violence to an extreme and make gruesome and exploitive brutality a selling point. For instance, games such as *Manhunt 2*, *Postal 2*, *Reservoir Dogs*, and *No More Heroes* have been criticized for containing grotesque torture, dismemberment, and other forms of sadistic and over-the-top violence (Young & Whitty, 2011).

Amidst all of the controversy that surrounds video games, the circumstances that surround the 2005 PC version of *Grand Theft Auto: San Andreas* are unique. This game contained senseless violence and criminality that are trademarks of the *Grand Theft Auto* franchise, but they were not the reason that *Grand Theft Auto*: *San Andreas* was, after release, re-rated to North America's most strict video game rating of *A* for *Adults Only* and withdrawn from the shelves of most major retail stores. Rather, this scandal was caused by an in-game mini-game that was deleted by the developers before release but then dug out of the game's code by the video game community. Dubbed the "hot coffee modification" the original version of this mini-game[1] allowed players to control their character to engage in fully clothed, poorly animated, and entirely consensual sex.

As with the quote that opened this chapter, the anecdote of the hot coffee modification serves to highlight a larger point; that there seems to be a discrepancy in the extent to which different types of fictional acts are condemned. More specifically, fictional sex seems to be judged more harshly than fictional violence. This is a relatively unexplored phenomenon within the

[1] The hot coffee modification was itself heavily modified by the gaming community after its initial discovery and through downloads, players were eventually able to engage in acts that were much more explicit that the developer's initial intent.

psychological literature, but this evaluative asymmetry between fictional sex and violence has been commented on in a number of different contexts. For instance, this discrepancy has been raised a in legal proceeding in which a US Supreme Court justice, in support of restricting violent video game sales, commented on the notable double standard between depictions of violence and sex, arguing that it was ridiculous to prohibit adolescents access to depictions of nudity, but allow access to depictions of violence (*Brown v. Entertainment Merchants Association, 2011*). In a similar vein, Leone (2004) systematically analyzed film rating criteria and came to the conclusion that the Motion Picture Association of America (MPAA) is more restrictive towards the depiction of sex than of violence within films. As a consequence of these contents being weighted differently, concerns have been raised regarding the extent to which a standardized rating scheme can actually allow consumers and parents to know if a media product will actually be appropriate for the designated age range (Thompson & Yokota, 2004). Lastly, the evaluative discrepancy between fictional sex and violence has been commented on in ethical philosophy by Luck (2009), who presents several normative arguments on whether people should ethically abide by society's double standard that frames engagement with violent media as more appropriate than engagement with sexual media.

This phenomenon will be labeled the *fictive pass asymmetry* because it seems that there is an evaluative asymmetry between the extent to which fictional depictions of violent versus sexual behaviors are morally permissible. The work that is presented in this thesis will, for the first time to my knowledge, offer systematic evidence of its existence and also

demonstrate the extent to which it is associated with moral judgments and emotions. As was mentioned earlier, when immoral behaviors occur in fictional contexts, it bears to reason that they will be evaluated less negatively than if they had occurred in the context of real life. In other words, fictional transgressions should get a “pass” so far as moral condemnation is concerned. However, as is made evident by the preceding discussion, there seems to be dissociation between the extent to which people are willing to tolerate fiction that is of a violent relative to a sexual nature.

More specifically, the present research will compare real acts to fictional acts. It will be demonstrated how a pass is given to fiction that encourages people to imagine immoral acts that cause harm to other people despite the same acts being wrong when they occur in real life. By comparison, fiction that encourages its consumers to imagine acts that are not obviously harmful, but violate moral norms of purity by involving an abnormal and counter-normative use of the body, will be given less of a pass and their condemnation will be closer to that of their real life counterparts.

In the moral psychology literature, harm code violations and purity code violations (Chakroff, Dungan, & Young, 2013; Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011; Rozin, et al., 1999; Russell, Piazza, & Giner-Sorolla, 2013) are most closely analogous to violent and sexual acts, respectively. Harm code violations involve harm being caused to specific individuals that can be in the form of physical violence, verbal degradation, material deprivation, or any sort of behavior that infringes upon one’s individual rights. Within the above literature, purity code violations have been variously defined as foul, disgust-evoking acts that can lead to literal infection

and contagion or symbolic corruption of one's soul or mind. The purity domain is a nebulous concept that will be discussed more thoroughly in chapter 2. For now, it will suffice to say that the work that is presented in this thesis will be taking the stance that purity code violations are most clearly defined as acts that violate bodily norms (Giner-Sorolla, Bosson, Caswell, & Hettinger, 2013; Russell & Giner-Sorolla, 2013). Acts that violate body norms may include, but are not limited to, strange sexual acts, counter-normative food consumption, or bizarre, but consensual, body-modifications like scarification or tongue splitting that many would see as an act of self-defilement.

To reiterate, the fictive pass asymmetry is interested in testing the extent to which the negative moral judgments and emotional reactions that are evoked from harm and purity code violations will carry over from real to fictional contexts. The most prominent literature that gives theoretical context to the fictive pass asymmetry hypothesis is that which demonstrates how harm and purity code violations give rise to different types of moral condemnation. Most notably, the person-centered approach to moral judgments (Uhlmann, Pizarro, & Diermeier, 2015) argues that some kinds of acts are immoral in and of themselves while others serve as indicators of poor moral character. More specifically, Chakroff and Young (2015) demonstrated how on the one hand, people explain their condemnation of harm code violations with act-based explanations that highlight the wrongness of the act itself; while on the other hand, acts that violated purity norms were more strongly attributed to person-based explanations. That is to say, the wrongness of the behavior was seen as going beyond the act itself

to signal information about the moral character of the one who committed it. Regarding the fictive pass asymmetry hypothesis, immoral acts that occur in fictional contexts do not cause any immediate harm and thus cannot be rationally condemned for their consequences. If, however, the fictional act is of an impure nature, it may still be condemned because of the extent to which it reveals information about the moral character of the one who chose to engage with the fiction.

Also, critical to the fictive pass asymmetry hypothesis are the moral emotions of anger and disgust. The boundaries between these emotions are not always clear and they regularly occur alongside one another (Russell, et al., 2013), but when they are distinct from one another, anger is more strongly associated with harm violations and disgust is more strongly associated with purity code violations. Anger is a flexible emotion that is most commonly evoked from perceptions of harmfulness such as physical aggression or injustice (Goldberg, Lerner, & Tetlock, 1999). Moreover, it is a flexible emotion that can be mitigated by the social context that surrounds the harmful act (Russell & Giner-Sorolla, 2011a; 2011b). In contrast, disgust is most strongly evoked from violations of purity and it has been characterized as a stubborn and inflexible emotion that responds more strongly to appearances than to reason (Rozin, Millman, & Nemeroff, 1986; Russell & Giner-Sorolla, 2011b). Based on how these two emotions function in the context of real life, the work that is presented in this thesis aims to demonstrate how the elicitation of anger that is felt towards immoral acts should drop substantially between real and fictional contexts. In contrast to this, the amount of disgust that is evoked from immoral acts should be

relatively more equal between real and fictional contexts because of its inflexible nature.

To summarize the hypotheses of this research, the fictive pass asymmetry hypothesis predicts that fictional acts of harm will be given a pass, meaning that they will evoke less moral condemnation and less negative emotion that their real-life counterpart. In contrast to this, purity code violations should be denied a pass regardless of the context in which they occur. The clearest expression of this would manifest as fictional purity code violations being subjected to just as much condemnation as their real-life counterparts, but it is possible that a drop between these contexts will still be present. Regardless, this evaluative discrepancy between reality and fiction should be smaller for purity code violations than for harm code violations, in relative, if not absolute, terms.

The research in this thesis will explore the fictive pass asymmetry hypothesis over six experiments that test the effect of different fictional contexts on moral judgments and moral emotions towards acts that violate harm and purity norms. For each experiment, evaluations of real acts were compared to evaluations of the same acts when they occurred in fictional contexts such as in imagination, as something that was watched in a film, or as something that was performed in a video game. Seeing as the focus of this research was on the extent to which moral judgments and emotions would cross the line from reality to fiction, there were no specific hypotheses regarding the differences between the fictional contexts, and for all experiments, they were aggregated into a single level of fiction. Nonetheless, chapter 7 does contain meta-analyses of select experiments from this thesis

and one of these analyses explores the relative size of fictive pass effects between the specific fictional contexts.

To give an overview of the chapters, the second and third chapters explore the literature that is most relevant to the fictive pass asymmetry hypothesis. Chapter 2 will discuss various components of moral psychology as they relate to the fictive pass asymmetry. This includes the development and the validation of the moral foundations theory (Haidt & Joseph, 2004), a more specific discussion of the importance of the harm and purity domains and their relations to moral judgments, and the chapter will close with a discussion of an alternate perspective on moral domains (Gray, Schein, & Ward, 2014) and how this alternate framework can be addressed by the fictive pass asymmetry hypothesis. Chapter 3 contains a more comprehensive discussion of the emotions of anger and disgust. Firstly, it will be explained how these emotions are distinct from another at their core, but they also share many similarities, especially in a moral context. Secondly, there will be a review of disgust's status as a moral emotion and it will be described how disgust's role in the moral realm, compared to anger, is most evident when it is felt in response to acts that violate bodily norms (Russell & Giner-Sorolla, 2013). Lastly, the varying flexibilities of anger and disgust will be explained in order to more clearly show how anger, as a flexible emotion, should show fictive pass asymmetry effects more clearly than disgust.

Chapters 4 through 6 describe six experiments and a series of meta-analyses that lend support to the fictive pass asymmetry hypothesis by demonstrating the extent to which moral judgments and emotions cross the line from reality to fiction. Chapter 4 reports three experiments that were

initial tests of the fictive pass asymmetry hypothesis. These three experiments show the systematic development of the methodology that allowed the effects of the asymmetry to most clearly present themselves. After the initial experiments succeeded in demonstrating the effects of the fictive pass asymmetry, those in chapter 5 will expand upon the experiments of the previous chapter in order to seek an explanation of the asymmetry. The two experiments that are reported in chapter 5 suggest that the asymmetry can be explained by the fact that purity code violations, more so than harm code violations, signal poor moral character. Chapter 6 firstly explores alternate explanations of the asymmetry by considering it in relation to a theoretical framework that challenges the distinctness of the moral domains (Grey et al., 2014). Secondly, chapter six presents the results of a series of meta-analyses that demonstrate the strength of the fictive pass asymmetry effects across the experiments.

In closing, chapter 7 will conclude the thesis by first discussing the six experiments and the meta-analyses in their entirety. Secondly, it will be discussed how the results of these experiments shed light on moral theories of act and character and on anger and disgust research. Lastly, limitations will be discussed alongside further research ideas and it will be discussed how the results of these experiments can inform on the applied role that moral psychology can play in regards to media use and regulation.

CHAPTER 2

MORALITY AND MORAL CONDEMNATION

This chapter will outline the psychological literature that has contributed towards the identification of distinct moral domains and explain the extent to which it has influenced the research that is presented in this thesis. First, the moral foundations theory (Graham et al., 2011) will be summarized. In addition to a discussion of the moral foundations theory itself, this section of the chapter will describe the theories and hypotheses that preceded the moral foundations theory, as well as more contemporary literature that offers it empirical support. Secondly, it will be explained how amongst the six moral domains of the moral foundations theory, the harm and purity domains are most directly relevant to the fictive pass asymmetry hypothesis. Specifically, it will be discussed how immoral acts that violate harm or purity norms have been shown to result in unique types of moral condemnation. To my knowledge, however, these patterns have only ever been demonstrated in the context of real life. To conclude this chapter, it will be emphasized how the type of condemnation that is evoked from harm and purity code violations in the context of real life should be able to explain variability in moral condemnation across fictional contexts, or in other words, the predicted effects of the fictive pass asymmetry.

Morality as conceptualized by the moral foundations theory

Morality is a large and complex topic that can help to explain core aspects of human nature such as our attitudes, beliefs, and behaviors.

Despite the expansive nature of morality, some believe that the breadth of the moral domain can be encapsulated with a single moral concern. For instance, Kohlberg (1971) argued that justice is the most essential element in understanding morality, and more recently harm concerns have been put forth as the foundation of all morality (Gray, Young, & Waytz, 2010). These individuals, referred to as *monists* by Graham et al. (2011), attempt to distil morality to a single domain. They argue that all moral concerns stem from one single facet of morality, but not surprisingly, they struggle to find consensus in regard to what exactly this single facet is. Instead, monists have put forth various arguments in which they promote the aspect of morality that they claim is most critical to our conceptualization of the moral world.

On the other hand, there are *pluralists* (Graham et al., 2012) who are less willing to accept that morality can be reduced to a single basic element. Instead they argue that many facets work in conjunction with one another in order to foster our understanding of morality. At the forefront of the pluralist school of thought is the moral foundations theory (Graham, et al., 2012), which attempts to deliver a flexible model of morality that can account for a wide range of moral concerns without being confined by a single domain.

Currently, the moral foundations theory is at the helm of psychology's conceptualization of morality, but the field did not always take a multifaceted approach to the moral world. As explained by Graham et al. (2011), the moral psychology literature used to be dominated by a monist school of though and the predominant belief was that various aspects of morality could be assessed, but morality itself was conceptualized as relatively more one

dimensional than our contemporary understanding would lead one to believe. As such, there were a variety of scales and measures to assess morally relevant traits and skills such as psychopathy (Levenson, Kiehl, & Fitzpatrick, 1995), moral reasoning (Rest, Narvaez, Thoma, & Bebeau, 1999), and empathy (Davis, 1983). These scales may have been able to measure certain aspects of morality, but they all worked under the assumption that morality revolves solely around concerns about harming other people. This is not necessarily a controversial view – it is still held by contemporary researchers (e.g. Gray et al., 2010) – but it is largely derived from the work of Kohlberg (1971) and Gilligan (1982) whose theories argue that morality can be understood by the extent to how well individuals treat one another.

As psychology's understanding of morality progressed, however, additional domains were argued for. For instance, researchers pointed out that a harm-based morality was mostly studied in relation to the Western English speaking world and less so in other cultures (Shweder, Much, Mahapara, & Park, 1997) such as in India, where concerns of spiritual purity are also prominent. These purity concerns, it was later found, also resonated in the United States, especially in regards to consensual sexual acts such anal sex that are objectively harmless (Haidt & Hersh, 2001). In effect, this demonstrated how harm is not sufficient as a sole foundation of morality and that in order to get a more comprehensive understanding of the moral world, additional facets of morality needed to be explored.

A frontrunner in establishing a more comprehensive taxonomy of moral concerns was Shweder et al.'s (1997) "Big Three" of morality, in which they proposed that moral concerns can be mapped onto three distinct

morally relevant issues, or ethics, as the researchers called them: those of community, autonomy, and divinity. Community ethics were framed as moral obligations that one had to obey in order to uphold the social hierarchy of one's community. For instance, these ethics would be violated by disrespecting authority figures, failing to perform the duties that your family or social group expect of you, or acting in a way that is not in the best interest of one's group. Shweder et al. (1997) argued that autonomy ethics consist of people's obligation to respect other individuals' independence and basic rights. These ethics are violated when one's actions infringe upon another's autonomy by, for instance, cheating, stealing, lying, or causing physical harm. Lastly, divinity ethics are framed as one's obligation to uphold the natural laws of the universe. These ethics may have religious overtones, although not necessarily, for while divinity ethics may be infringed upon by directly disrespecting God, they can also be violated by performing more generally impure and degrading acts that do not explicitly intend to violate religious tenets.

In subsequent research, psychology's understanding of these three ethics was deepened when Rozin, Lowery, Imada, and Haidt (1999) proposed the CAD triad hypothesis. This framework drew specific links between Shweder and colleagues' (1997) moral codes of community, autonomy, and divinity with the emotions of contempt, anger, and disgust, respectively. The relationship between these emotions and these moral codes will be discussed in more detail in the following chapter, so for now the focus will be strictly on moral domains. The emotions aside, however, the CAD triad was highly influential in shaping the literature on moral domains

and while it has been criticized (Hutcherson & Gross, 2011; Russell, et al., 2013), it laid the groundwork for the development of the more comprehensive moral foundations theory.

The taxonomy of the moral realm was further developed by Haidt and Joseph (2004) who rebranded, and expanded upon, the three domains that were initially proposed by Shweder et al. (1997) and Rozin et al (1999). With the introduction of the moral foundations theory, Haidt and Joseph (2004) went beyond the existing categories in order to create a more comprehensive classification of moral values that they believed best encapsulate people's reactions to immoral behavior.

Using taste buds as an analogy to describe their initial iteration of the moral foundations theory, the researchers explained how different cultures' tastes for food differ greatly, yet all gastronomic pleasures, regardless of cultural variants, are derived from innate and universal taste receptors (Haidt & Joseph, 2004; Haidt & Graham, 2007). Similarly, cultures vary in the extent to which they weigh and understand moral intuitions and as such, one must always be wary of claiming universal truths about morality. Regardless, the moral foundations theory argues that there are five domains of morality that regularly reveal themselves around the world and have evolved alongside our respective cultures. Haidt and Graham (2007) list and explain the five domains; harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity.

1) The h*arm/care* foundation roughly parallels the autonomy ethics of the CAD Triad Hypothesis (Rozin et al., 1999) in that it

is based in a desire to minimize harm towards other individuals. The authors write that the average person has an aversion to harm and cruelty and as such, we praise those that care for others and condemn those that cause harm.

2) The f*airness/reciprocity* foundation also shares similarities with the CAD Triad Hypothesis' autonomy ethics in that these concerns are triggered by the perception of cheating and unfairness.

3) The *ingroup/loyalty foundation* is related to human's long history as social animals and it serves to regulate virtues such as patriotism and loyalty. It expands upon the CAD Triad Hypothesis' ethics of community by considering the horizontal nature of one's social obligations.

4) The *authority/respect* foundation is derived from our history of living in societies with hierarchically structured ingroups. It expands upon the community ethics of the CAD Triad Hypothesis by emphasizing the vertical nature of society by incorporating ideals such as leadership, authority, and respect.

5) The *purity/sanctity* foundation most closely resembles the divinity ethics of the CAD Triad Hypothesis in that it was shaped by concerns of disgust and contamination. Like its predecessor, this foundation does not presume religious beliefs; this foundation is nonetheless of a spiritual nature in that it promotes living a morally clean life. Critically, this requires that one does

not involve one's body in immoral or contaminating activities (Giner-Sorolla, et al., 2012; Russell & Giner-Sorolla, 2013).

A large body of work has been dedicated to validating these domains, demonstrating the broad application of the moral foundations theory, and the extent to which the foundations are associated with various social and political factors. In terms of political orientation, United States liberals and conservatives differ in the extent to which they endorse the different moral foundations. Liberals, more so than conservatives, are more concerned about issues that relate to the foundations of harm and fairness whereas conservatives, relative to liberals, are more concerned about issues that relate to the foundations of ingroup, authority, and purity (Graham, Haidt, & Nosek, 2009, Experiment 2). The Moral Foundations can also inform personality psychology, as demonstrated by Koleva, Selterman, Iyer, Ditto, and Graham (2013) who found that individuals' attachment styles can predict variance in their endorsement of different moral foundations. For instance, participants with high attachment anxiety, relative to those with low attachment anxiety, were more concerned with issues relating to harm and unfairness. As indicated by these examples, the potential uses of the moral foundations theory are broad and far reaching. Perusing its website (moralfoundations.org) reveals a tremendous array of publications that demonstrate how the moral foundations theory can inform on issues ranging from domestic abuse (Vecina, 2014), to sports fandom (Winegard, & Deaner, 2010), and even people's motivation to adapt an ecologically friendly lifestyle (Kidwell, Farmer, & Hardesty

(2013). The most apparent application of the foundations is, however, furthering the understanding of how one's moral intuitions relate to various socio-cultural ideologies. Clearly, there is a wide breadth of applications across all six foundations. Nonetheless, it is important to isolate and discuss two specific foundations that are most directly relevant to the fictive pass asymmetry hypothesis.

The Moral Domains of Harm and Purity

Of the five domains of the moral foundations theory, the domains of harm/care and purity/sanctity (harm and purity for the sake of concision and to be in line with much of the existing literature) are the most directly relevant to the research that is presented in this thesis. A fair amount of recent research – beyond that which has validated and explored the applications of the Moral Foundations Theory (e.g. Graham et al., 2009; 2011) – has focused more directly on demonstrating that harm and purity code violations are unique moral domains that are associated with specific types of behaviors and result in different types of evaluations.

It should be reiterated how acts that violate the harm domain may include physical acts of violence, material deprivation, or verbal abuse. Most critical is that these harm violations are perceived as resulting in specific negative outcomes for other people. Acts that violate the purity domain are most prominently acts that violate bodily moral norms (Giner-Sorolla, et al., 2013; Russell & Giner-Sorolla, 2013). Purity violations may include sexual acts that are seen as wrong and immoral regardless of consent (e.g. consensual incest), food consumption that violates religious laws, or acts of

bodily modification such as scarification or bizarre piercings that are seen as a defilement of the self. Furthermore, harm and purity code violations are most strongly associated with the moral emotions of anger and disgust, respectively. These emotions, however, and their links with these moral codes will be discussed more thoroughly in the next chapter that will focus specifically on emotions.

A number of experiments have been dedicated to specifically exploring the harm and purity domains in order to understand the types of acts that evoke these concerns and how people evaluate such behaviors. These experiments give theoretical context to the fictive pass asymmetry hypothesis by identifying unique ways in which harmful and impure acts are evaluated. First consider the research that displays the relative malleability of these two different types of violations. Evaluations of harm violations are flexible and thus susceptible to rationalization, justification, and social context (Gutierrez & Giner-Sorolla, 2007). Physical violence, for instance, may be seen as immoral until it is revealed that it was an act of self-defense. Similarly, theft is generally wrong, but less so when out of economic necessity or when the one being stolen from is unfairly hoarding resources. It is difficult, however, to similarly justify violations of purity. As Haidt (2001) described in his paper, an incestuous relationship is seen as wrong, regardless of the extent to which it is framed in a positive light. Evaluations of purity code violations are, by their very nature, associative, unreasoned, and thus highly resilient to mitigating circumstances (Giner-Sorolla, 2012; Russell & Giner-Sorolla, 2013). Consequently, even when a purity violation is entirely consensual and free of all harm, it will still evoke intuitively negative gut

reactions even when people are unable to elaborate on the reason for their condemnation (Haidt, Bjorklund, & Murphy, 2000; Haidt, Koller, & Dias, 1993; Uhlmann & Zhu, 2013).

Moreover, researchers have begun to identify an act-character divide in moral judgments of harmful versus impure acts (Pizarro, Tannenbaum, & Uhlmann, 2012; Tannenbaum, Uhlmann, & Diermeier, 2011; Uhlmann, Zhu, & Diermeier, 2014). To contextualize the literature that has established these dissociations between the focus of one's moral condemnation, consider the public perceptions of Michael Vick, an NFL quarterback. In 2007, Vick was accused of hosting, funding, and organizing an interstate dog fighting league. After these repugnant activities came to light he was greatly vilified, and journalist Roland Martin (Martin, 2007 as cited by Uhlmann, et al., 2015) claimed that Vick's public image would have been less tarnished if he had instead been accused of beating his girlfriend or even murdering another human being. Rationally, it is odd that one would be condemned more for harming an animal over another human being, but this anecdote highlights a larger point: some acts are bad in and of themselves but others may reveal unsavory information about one's character.

In an initial exploration of this phenomenon, Tannenbaum et al. (2011, Experiment 1a and Experiment 1b) demonstrated how certain types of acts prompt character-based evaluations over and above the condemnation of the act itself. Violence towards a human was seen as more immoral than violence towards a pet cat, but violence towards the cat was seen as a signal that one lacks empathy and thus served to signal substantial deficits in the offender's moral character. In their final experiment, Tannenbaum et al.

(2011, Experiment 2) generalized these findings by showing how a similar asymmetry can form even in the absence of physical harm. In the task that was presented as part of the experiment, participants opted to hire an extremely well paid CEO over an equally qualified candidate who would cost the company less money. The lesser paid, but equally qualified candidate did, however, request that he was given the frivolous perk of having a self-portrait engraved in his desk. In addition to not being hired, participants reported that the second candidate had less integrity than the former, suggesting that – as with the findings involving violence towards the pet cat – strange behaviors that lack any sensible consequential end may stem from untrustworthy or sadistic character traits.

It is of course acknowledged that Tannenbaum and colleagues' experiment does not directly inform on the differences between harm and purity code violations. Instead, it explores the evaluative consequences of harmful (both physical and financial) acts versus ones that are relatively less harmful, but substantially more bizarre. An objectively harmless purity code violation such as Haidt's (2001) example of consensual sibling incest can also fit this criterion of an act that is harmless but bizarre. It is, therefore, reasonable to predict that a harmless act of impurity, relative to a more prototypically harmful act, would evoke more character than act-based condemnation.

Before a discussion along this line of thought continues, it is important to have a brief digression to discuss moral dumbfounding (Haidt, et al., 2000, Haidt, 2001). The principles of moral dumbfounding heavily influenced some of the research that followed Tannenbaum et al.'s (2011) experiment and

helped to draw a distinction between act and character-based judgments via the gut feeling of “wrongness” that is moral dumbfounding.

The psychological phenomenon of moral dumbfounding plays an important role in understanding why impure acts, specifically those that are objectively harmless, are still subjected to moral condemnation. To study moral dumbfounding, researchers have typically presented participants with scenarios that depict harmless, but impure acts. These may include acts such as having sex with a dead chicken carcass (Uhlmann & Zhu, 2013) or engaging in consensual sibling incest (Haidt, Björkland, & Murphy 2000). In the descriptions of these acts, researchers make a deliberate effort to highlight the entirely harm-free nature of these acts (i.e. safe, private, no psychological trauma or long term damage). No matter the lengths that researchers go to while crafting these vignettes, participants are still inclined to judge the behaviors that they describe as morally wrong. Despite participants’ condemnation, however, they struggle to provide elaborate justifications for their judgments and they instead rely on trivial reasoning; condemning the act for the simple fact that it is wrong, or justifying their condemnation based on the fact that makes them feel disgusted.

Indeed, research on moral dumbfounding (Haidt et al., 1993; Haidt et al., 2000; Uhlmann & Zhu, 2013) suggests that all one needs in order to condemn impure behaviors are the tautological reasons of basic judgments alongside the emotional reactions of disgust. This perspective on morality, known as the social intuitionist account of moral judgments (Haidt et al., 2000), argues that judgments are based largely on gut reactions, and that reasoning serves primarily to construct post-hoc rationalizations of these

reactions. The emotion of disgust is a critical aspect of people's intuitive reactions to impure behavior (Russell & Giner-Sorolla, 2011a), but the role of this emotion will be discussed more in the next chapter.

To continue, the previously discussed principles of moral dumbfounding will now be used to further illustrate the divide in act and character-based moral judgments that was demonstrated by Tannenbaum et al., (2011). This notion was initially proposed by Uhlmann and Zhu (2013), who argued that moral dumbfounding could be better understood if it were examined in relation to the act-person distinction in moral condemnation. Across their experiments, their general procedure involved presenting participants with a description of a man committing either a harmless purity code violation or a blatant act of harm. Further randomization then asked participants to make judgments about the main character's act itself or about his moral character and the extent to which they could confidently explain their condemnation. The results indicated that participants became morally dumbfounded when they were asked to provide act-based explanations for their condemnation of harmless purity violations. By contrast, when participants were asked to provide character-based explanations of their condemnation of these acts, they were relatively less morally dumbfounded. The authors argue that this is because impure acts offer a goldmine of information regarding the state of one's character, leaving participants with a clear understanding of their condemnation.

Continuing this line of thought, Chakroff and Young (2015) offered a more explicit comparison between harm and purity code violations by demonstrating an act versus character-based attributional asymmetry

between these two types of acts. In their experiments, they presented participants with acts that violated harm norms (described as various acts of physical violence) and purity norms (described as acts such as incest or handling fecal matter). The results indicated that impure acts, more so than harmful acts, were attributed to character-based explanations rather than act-based explanations. By contrast, participants more strongly endorsed act-based explanations for harmful acts, despite the fact that they perceived both types of acts to be equal in moral wrongness. Effectively, this series of experiments offered more concrete evidence that impure acts, relative to harmful acts, are more indicative of bad character.

The Importance of Moral Character

Moral character, as will be demonstrated in chapter 6, is a critical part of the fictive pass asymmetry hypothesis. Goodwin, Piazza, and Rozin (2014) write that moral character consists of the moral dimension of one's personality and that it is an important element in understanding other people and navigating the social world. The ability to intuit the state of another person's moral character is a valuable social skill to have because it allows one to gauge the extent to which another individual will be either helpful and cooperative or harmful and antisocial (Abele & Wojciszke, 2007; Hogan, 1973). More generally, moral character is a key process that underlies the impression that we initially form of other individuals (Wojciszke, Bazinska, & Jaworski, 1998). In spite of the important role that moral character plays in people's daily lives, Goodwin and colleagues (2014) claim that most

research on the topic is fundamentally flawed because character is conflated with other dimensions of personality, and thus it frequently suffers from poor measurement.

For instance, in an early model of person perception, Rosenberg, Nelson, and Vivekananthan (1968) instructed participants to sort a vast number of personality traits based on the dimensions of good-bad, hard-soft, and active-passive. Ultimately, the traits were fitted onto two axes that categorized them based on the extent to which they were perceived as good versus bad and social versus intellectual. While many of the traits on these dimensions were not necessarily related to morality (i.e. skillful, squeamish, popular) many of them were (i.e. honesty, warmth, sincerity), yet these traits that seems to capture aspects of one's moral character were amalgamated into axes that may have inappropriately framed character as a homogeneous construct.

This mindset has been fairly evident throughout some of the more recent literature. Perhaps the most well-known successor of Rosenberg et al.'s (1968) dual axis model is the more contemporary stereotype content model (Cuddy, Fiske, & Glick, 2008). This model posits person perception and evaluation as being composed of the two independent traits of warmth and competence, and moral character is placed on the axis of warmth. As with its predecessor, this homogenization of character prevents researchers from fully understanding the extent to which it may be a multifaceted construct and stems future research from exploring the specific traits that it may be comprised of.

Take, for instance, the aforementioned literature that has attempted to distinguish between act and character-based moral judgments (i.e. Pizarro, et al., 2012; Tannenbaum, et al., 2011; Uhlmann, Zhu, & Diermeier, 2014). Despite successfully differentiating between these two types of moral judgments, the experiments presented in these papers assessed character in a general, and perhaps oversimplified, way. For instance, Uhlmann and Zhu (2014, Experiment 2a) explored the extent to which harmful versus harmless-but-impure acts would elicit greater levels of act or character-based condemnation by randomly presenting participants with a scenario of a man stealing food or a man eating a dead dog that had been killed by a car. Act and character condemnation were measured with single Likert-type items that asked, “Is this morally wrong?” and “Does this person have poor moral character?” Similarly, Chakroff et al. (2013, Experiment 1) assessed the moral character of an individual described in a vignette with a Likert-type item that asked “How immoral is Steve?”

Despite the evident shortcomings of amalgamating the components of moral character, the research that is presented in this thesis does, more often than not, follow the lead of the most prominent research in the field. As such, the construct of character has been measured in a way that is less comprehensive and complex than what is recommended by Goodwin et al., (2014). Chapter 6’s Experiment 6 does, however, more directly account for the advice of Goodwin and colleagues by acknowledging that moral character is not a homogenous construct and the measures more thoroughly account for its multi-component nature.

In sum, there are distinct differences between the moral codes of harm and purity and the extent to which they elicit act-based or character-based condemnation. Evaluations of harm code violations, to a lesser extent than purity code violations, are susceptible to contextual factors (Gutierrez & Giner-Sorolla, 2007) and offer less information about one's moral character (Uhlmann & Zhu, 2013; Chakroff & Young, 2015). In relation to the fictive pass asymmetry hypothesis, acts that are engaged within a fictional context do not cause any clear harm to specific individuals. They may, however, reveal information about the status of one's moral character and it is even possible that they are seen as corruptive forces that have the ability to turn one into an immoral person. In the context of real life, acts that harm other people (i.e. harm code violations) have obvious consequences and may, therefore, be condemned more than immoral acts that do not cause harm (i.e. objectively harmless impurities). To experience either of these acts in fiction, however, demonstrates that one toys with the idea of committing morally blameworthy or antisocial acts. Due to its lack of consequences, experiencing the act in fiction should result in negligible act-based judgments. The character-based judgments should, however, be relatively more similar between real and fictional contexts, because in fiction the lack of consequences means that the act itself cannot be justifiably condemned, thus leaving moral character as the scapegoat.

From the preceding discussion, it has hopefully been made clear that the research in this thesis has built its foundation on the notion that harmful and impure acts form distinct moral domains. The distinctness of this partition is not, however, a point of view that is entirely agreed upon in the

moral psychology literature and some researchers instead argue for a *lack* of distinctness between harm and purity, taking the stance that they are one and the same.

An Alternate Perspective on the Harm – Purity Distinction

This alternate view on these two different immoral acts describes both harm and purity code violations, even purity violations that are objectively harmless, as involving harm, only to different targets (Gray, et al., 2014). In line with the moral foundations theory, Gray et al., (2014) agree that acts that violate the commonly called “harm” domain involve injury to specific others. Contrary to the moral foundations theory, however, Gray and colleagues argue that acts typically classified as violating the “purity” domain also cause harm, just to non-specific entities such as society, God, memory, or one’s spirit.

It is argued that the distinction between harm and purity code violations as described in the moral foundations theory does not accurately describe a layperson’s understanding or interpretation of immoral behavior. Harm is perceived in seemingly innocuous violations of purity because individuals have a need to complete a template of morality (Gray & Wegner, 2010; Gray, Waytz, & Young, 2012; Gray, et al., 2014; Schein & Gray, 2016). According to Gray, Young, and Waytz (2012), an individual’s moral template is dyadic in nature and in order for one to understand and interpret immorality, such acts require both a victim and a perpetrator. In the absence of a clear victim, however, Gray and colleagues argue that one’s mind will intuitively infer the presence of a victim – and in turn perceive an act of harm

– in order to successfully satisfy one's dyadic template of morality. In effect, this process makes it so that wrong acts are perceived as harmful and vice versa (Schein & Gray, 2016), thus allowing individuals to infer that all immoral acts are harmful to some extent because if they do not cause specific harm to another person, then they harm an abstract entity such as God or the natural order of things.

This stance of a strictly harm-based morality works in opposition to the fictive pass asymmetry hypothesis, which – as with much other research – focuses on harmless but abnormal acts that are seen as immoral, which can be defined as objectively harmless purity code violations or, perhaps, as harm towards unspecified entities. Experiment 6, which is presented in chapter 6, will attempt to address and resolve this conflicting framework by testing whether the predicted fictive pass asymmetry also occurs between acts that do, versus do not, cause harm to specific entities. However, throughout the majority of this thesis, and in line with much of the existing literature on moral violations, immoral acts that directly harm other people and immoral acts that evoke disgust and involve the body yet do not harm specific others, will be referred to as harm and purity code violations, respectively.

Conclusion

In closing, the recent literature on moral psychology has developed taxonomy of moral domains. Most recently, the moral foundations theory (Graham et al., 2012) puts forth the notion that there are six domains of morality than can help researchers explain and understand the origins and

evolution of people's moral reasoning. Of these six domains, those of harm and purity will be explicitly explored in the work that is presented in this thesis. These two domains will be used to demonstrate the hypothesized effects of the fictive pass asymmetry hypothesis, which predicts that fictitious harm that is directed towards other people will be subjected to less moral condemnation than the same act of harm that is committed in the context of real life. In contrast to this, fictitious purity code violations should be denied a pass because unlike harm, impurities are condemned because they are more strongly associated with person-based attributions. When an act of harm is committed in the context of real life, it will be condemned because of the extent to which it is perceived as being harmful to other people. In fiction, where there is no one to harm, this condemnation should effectively dissipate. In fiction, however, both types of acts suggest that one has engaged with ideas of committing moral transgressions and this should lead to more equivalent evaluations of negative character.

CHAPTER 3

ANGER AND DISGUST

The current chapter will present a discussion of the moral emotions of anger and disgust. Intuitively, one would most likely say that these emotions are distinct in that they are elicited from different stimuli or situations and that they are accompanied by unique subjective feelings. A substantial body of literature does, however, suggest that this is not always the case. Anger and disgust can be used synonymously with one another and they can even be confused with one another. This is especially true when these emotions are being expressed towards moral transgressions. When they are distinct from one another, however, the extent to which these emotions are experienced towards the enactment of fictional code violations should provide additional evidence for the fictive pass asymmetry hypothesis.

The self-conscious and other-condemning moral emotions

The last chapter raised, but did not thoroughly discuss, how the social intuitionist model of emotion proposes that moral judgments are primarily driven by gut feelings and that rationalization and reasoning occur post-hoc in order to justify one's intuitive reaction to the morally relevant behavior (Haidt, 2001). Haidt (2001) argues that these gut feelings are driven by moral emotions, or emotions that functionally motivate and encourage moral behavior and decision making (Giner-Sorolla, 2012; Haidt, 2001; 2003). As said by Tangney, Stuewig and Mashek (2007), "moral emotions provide the motivational force – the power and energy – to do good and to avoid doing

bad" (p. 346). Moral emotions can be broadly divided into two clusters: the self-conscious emotions such as shame, guilt, and embarrassment and the other-condemning emotions such as anger, disgust, and contempt (Tangney et al., 2007).

Both the self-conscious and other-condemning emotions play a large role in people's understanding of morality. Of these two clusters of emotions, however, the self-conscious emotions were not included as part of the research that is presented in this thesis. Similarly, contempt, despite its status as an other-condemning emotion, was also excluded from this research in favor of focusing most directly on anger and disgust. Although contempt was not directly examined in this research, it does not deserve to be entirely disregarded because it still shares close ties with anger and disgust, as indicated by the name of Rozin et al.'s (1999) CAD (contempt, anger, and disgust) triad hypothesis. As discussed in the previous chapter, Rozin and colleagues (1999) established links between these three emotions and the moral codes that they categorized as community (e.g. disrespecting the social hierarchy), autonomy (e.g. physical harm), and divinity (e.g. counter-normative sexuality). This chapter will primarily focus on contrasting and comparing anger and disgust but because of contempt's ties to these other emotions it will be briefly discussed at this stage.

Contempt is a challenging emotion for emotion researchers, especially those who focus on the other-condemning emotions. One way in which it has proven to be challenging is in the fact that it shares similarities with both anger and disgust, suggesting that it might be, in a sense, a combination of these emotions. For instance, contempt is similar to anger in

that it can involve blame, verbal degradation, and even anger itself (Fischer & Roseman, 2007). At the same time, however, it has been demonstrated how contempt can be more akin to disgust in that its result action tendencies can also be “cold” in that they may include rejection, avoidance, and social exclusion (Fischer & Roseman; 2007; Roseman, Wiest, & Schwartz, 1994). Of course, these consequences of contempt only scratch the surface of an emotion that is fraught with theoretical and methodological challenges (see Fischer & Giner-Sorolla, 2016). These challenges do, however, demonstrate why the emotion of contempt has been excluded from the present research in the sake of focusing on the relatively more clearly defined emotions of anger and disgust. These two emotions are highly relevant to the research that is presented in this thesis and their similarities, differences, and their relevance to the fictive pass asymmetry hypothesis will now be discussed in more detail.

The similarities and differences of anger and disgust

The emotions of anger and disgust have received a lot of attention in recent psychological research. From a layperson’s perspective, one would assume that people know what these words mean and what these emotions feel like. Moreover, one would assume that these emotions are used in different contexts and that they are distinct from one another. In support of this intuitive perspective on these emotions, experimental work has reliably demonstrated that these emotions are accompanied by distinct and recognizable facial expressions, behavioral outcomes, and physiological responses. To illustrate these three primary differences, Eckman (1992;

1999) argued that anger and disgust, owing to their unique facial expressions, are two of the six basic emotions – although they both involve a raised upper lip and are thus confused for one another more consistently than the other basic emotions (Giner-Sorolla, 2012). Secondly, Roseman, Wiest, and Swartz (1994), asked participants to recall life experiences that were associated with strong emotional reactions and they found that the emotions of anger and disgust were associated with different action tendencies. Disgust was associated with withdrawal tendencies whereas anger was associated with more hostile and approach oriented action tendencies (also see Gutierrez and Giner-Sorolla (2007, Experiment 2). Lastly, from a physiological standpoint, anger causes one's heart to race, but disgust causes one's heart rate to slow down (Levenson & Ekman, 2002).

To recap, a portion of the literature on anger and disgust has reliably demonstrated that anger and disgust are distinct and unique emotions that vary in a number of substantial ways and possess many unique features. In spite of these differences, however, it is not uncommon for them to be uttered in the same breath, and even used interchangeably. Giner-Sorolla (2012) illustrates this linguistic similarity by sharing a number of amusing instances in which irate politicians, bloggers, and popular Twitter accounts have expressed the extreme nature of their disapproval by saying how they are both angry and disgusted simultaneously. Indeed, empirical approaches have also examined similarities between these two emotions. Such experiments have demonstrated that in spite of the clear differences between these emotions, anger and disgust, as well as a number of their synonyms (i.e. rage, revulsion) were shown to have clear semantic links (Shaver,

Schwartz, Kirson & O'Connor, 1987). Similarly, Russell and Fehr (1994) discuss how emotion words, especially those for anger, are a nebulous and fuzzy concept. Through a variety of methodologies across seven experiments the authors highlight a substantial overlap between the language of anger and disgust (amongst other emotions). For instance, participants in Experiment 5 indicated that disgust was a fair substitute for anger and in Experiment 7, 69% of participants went so far as to categorize disgust as a subcategory of anger (Russell & Fehr, 1994).

In sum, a large body of literature indicates that anger and disgust are distinct emotions that have unique facial expressions, physiological responses, and action tendencies. Challenging this, however, is another body of work that emphasizes the similarities, especially the linguistic similarities between these two emotion words.

In light of this confusion, it is important to understand the root of these conflicts and to establish a clear way in which these two emotions can be peeled apart. To assist in this disambiguation, Giner-Sorolla (2012) suggests that one engages in a thought experiment in which one is the lone inhabitant of a desert island because in the absence of other people and when one is focusing strictly on survival, the differences between anger and disgust are most evident. Anger would be directed towards things that block goals such as strong winds that blow down your shelter, a monkey that steals your food supply, or the vines that are unable to hold together your makeshift raft. The elicitation of disgust, on the other hand, would be exclusively related to biological concerns of infection or contamination. For instance, disgust would be evoked from things such as rotten meat and bodily waste. To elaborate

on the way in which disgust is elicited on this island, consider the extent to which we are surrounded by pathogens and the extent to which we are motivated to avoid them. Pathogens are absolutely everywhere, and across the species a wide variety of strategies have been adapted in order to help avoid infectious substances, potential pathogens, and other sources of contamination. Elephants, for instance, create makeshift flyswatters out of small leafy branches in order to keep at bay flies that could be carrying blood borne pathogens (Hart, 2011) and similarly, cattle reduce the risk of intestinal parasites by avoiding grass that is adjacent to their droppings (Dohi, Yamada, & Entsu, 1991). For humans, our most prominent method of disease avoidance is the emotion of disgust. The type of disgust that functions as a disease avoidance mechanism, and is brought on by potential sources of contamination, is referred to as core disgust (Curtis & Biran, 2001; Oaten, Stevenson, & Case, 2009). When comparing anger to core disgust, the differences are quite clear, and this disparity is made most clear when you contextualize these emotions in an environment, like the desert island, that is both non-social and non-moral.

Disgust as a moral emotion

The boundaries between anger and disgust do, however, become hazy when other people are introduced into the equation and these emotions take on a socio-moral component (Haidt, 2003). To illustrate this similarity, imagine a small business owner who embezzles money, gets caught, and consequently causes his employees to lose their livelihoods. A common reaction to this news might be saying that the business owner and his

actions are disgusting, despite the fact that they were physically sanitary and not at all related to pathogen concerns. It is just as likely, however, that people would report feeling angry because the business owner's greed caused innocent employees to lose their jobs. Effectively, in a socio-moral context, anger and disgust can fill a similar role and disgust can be used as an anger synonym to strengthen the expression of one's moral disapproval (Gutierrez, Giner-Sorolla, Vasiljevic, 2012). This unique aspect of disgust can be explained by the evolutionary term of exaptation, a process that allows adaptive traits to maintain their original purposes of avoiding physical contaminants while simultaneously serving more contemporary, and in this case social purposes, such as avoiding moral impurities.

This notion that moral disgust is an extension of pathogen concerns can be traced back to the research of Rozin, and colleagues (1986) who demonstrated that the elicitation of both physical disgust and moral disgust follows the irrational principles of sympathetic magic, a branch of mysticism that is based of imitation and contact. For example, one of their experiments demonstrated that when a dead and thoroughly sterilized cockroach was put into a glass of juice, participants were no longer interested in drinking the juice. Granted, this finding is entirely unsurprising, but more surprising is that participants' disinterest spread to an entirely different cup of that same type of juice. On one hand, this demonstrates the unwavering effectiveness of disgust as a pathogen avoidance mechanism. On the other hand, and more critical to the discussion on moral disgust, was another experimental condition that demonstrated similar principles of "contagion" when participants showed an unwillingness to wear a laundered shirt that allegedly

belonged to someone they disliked. This is interesting because unlike in the cockroach example, there were no traces of pathogen disgust whatsoever, indicating that the unpleasantness of the shirt must have stemmed from the character of its purported owner, suggesting that the avoidant action tendencies that are inspired by physically disgusting objects can also be prompted by morally impure individuals.

Four positions on moral disgust

This above example demonstrates the contagious aspects of moral disgust, but it does little to actually disambiguate anger and disgust in a moral context and showcase them as distinct emotions. Russell & Giner-Sorolla (2013) discuss four different stances that describe the specific role that disgust, relative to anger, plays in a moral context. The authors classify these four positions as the *general morality* position, the *purity* position, the *metaphorical use* position, and finally they propose the *bodily norm* position.

The general morality position argues that there is a general link between the emotion of disgust and moral condemnation. For instance, experiments have induced disgust in participants to demonstrate how it will enhance moral condemnation. One such experiment hypnotized participants to feel quick pangs of physical disgust when they read a target word. In the second half of the experiment, participants read and evaluated a series of vignettes, some of which contained the trigger word. These vignettes described a variety of immoral behaviors that ranged from sexual taboos and impure food consumption to cheating and bribery. For the hypnotized participants, the vignettes that contained their trigger word were condemned

more than those that did not contain the trigger word (Wheatley & Haidt, 2005). Similarly, it has also been shown that if disgust is orally induced through drinking foul tasting drinks, it will enhance more condemnation relative to a control group (Eskine, Kacinik, & Prinz, 2011).

Further to this, Hutcherson and Gross (2011) showed how a variety of socio-moral violations such as cheating, dishonesty, and embezzlement serve as disgust elicitors. Even more generally, Rozin, Haidt, and McCauley (1993, as cited in Russell & Giner-Sorolla, 2013) asked North American and Japanese participants to list disgusting things and they came up with a list that covered a variety of sexual and non-sexual violations, thus demonstrating a general link between disgust and immoral activity that is not necessarily restricted to specific types of behaviors.

In the main, the general morality position succeeds in demonstrating a clear link between disgust and immorality. This position, and the literature that supports it do not, however, successfully account for the possibility that anger, more so than disgust, is responsible for participants' condemnation. This is especially true in regards to the socio-moral violations that do not contain any elements of physical disgust. Thus, Russell and Giner-Sorolla (2013) describe the *purity* position, and offer a more specific definition of moral disgust in an attempt to disambiguate it from anger.

As discussed in the previous chapter, the CAD triad hypothesis and its successor the moral foundations theory (Rozin al., 1999; Haidt & Graham, 2007), mapped moral concerns onto various moral domains and these two models gave way to a more precise definition of moral disgust. Specifically,

these models demonstrated how disgust was most specifically evoked from behaviors that were described as being impure. It was argued that these disgusting acts, described as divinity violations by the CAD triad hypothesis and subsequently as purity violations by the moral foundations theory, were seen as behaviors that contaminated one's mental, physical, or spiritual purity.

For instance, Rozin and colleagues (1999) demonstrated a difference in the extent to which impure acts versus social-moral acts evoked different levels of emotion. Impure acts such as biting into an apple that has a worm inside it, touching a dead body, or eating rotten meat, all evoked more disgust than anger. On the other hand, socio-moral violations such as jumping the queue, stealing money from a blind person, or beating one's spouse, were all found to be associated more strongly with anger than with disgust. This experiment demonstrates how disgust is the most prominent emotional reaction to acts that are of a specifically impure nature, compared to the socio-moral violations that primarily evoke anger. Further to this, it has also been shown that feelings of disgust, but not anger, heighten the levels of moral condemnation that are directed towards impure acts (Horberg, Oveis, Keltner, Cohen, 2009).

The literature has shown that purity code violations are linked to disgust, to a greater degree than they are linked to anger. In spite of this, it is not satisfactory to claim that impurity is the sole element that separates these two emotions; often times, purity concerns are conflated with the bodily aspects of disgust. For example, consider the impure acts that were presented by Rozin et al., (1999) such as eating a worm or touching a dead

body. These acts are impure, but they also involve people using their bodies in an abnormal fashion. Consequently, it is unclear whether anger and disgust are most distinct when someone acts in a way that is impure, or when someone uses their body in a counter-normative way,

Russell & Giner-Sorolla (2013) discuss how this ambiguity is problematic considering that purity code violations, by definition, are not predicated on the immoral use of the body, but rather on principles of living a clean life in which one is pure of body, but also of mind and spirit. Nonetheless, much of the literature on purity code violations seems to ignore these latter facets of purity and instead focuses on its bodily aspect. As a result, it is not entirely clear if behaviors can be impure, contaminating, and disgusting, all without actually involving one's body. Admittedly, the research that is presented in this thesis does not attempt to fill this notable gap in the moral disgust literature. Instead, the principles of the fictive pass asymmetry hypothesis are based on the bodily-norm position of moral disgust (Russell & Giner-Sorolla, 2013); a position that will be discussed in more detail after a brief summary of the metaphorical use position.

Many of the most fundamental facets of the metaphorical use position have been discussed earlier in this chapter. For instance, it was described how in the moral realm, anger and disgust are semantically linked – at least in the English language -- when they are being directed towards socio-moral violations such as stealing, cheating, or racism and it is not uncommon for them to be used interchangeably (Shaver et al., 1987).

As a result of this overlap some researchers are of the opinion that, so far as morals are concerned, common disgust language theoretically encompasses both anger and disgust. For example, Nabi (2002) asked an undergraduate sample to recall and write about a time that they felt disgusted, revolted, “grossed out”, or angry. A qualitative analysis of participants’ responses suggested that their understanding of the term *disgust* closely reflected the theoretical definition of anger in that the words “disgust” and “disgusted” were used to describe infringements on people’s autonomy such as disparity, cheating, and disrespect. The term “grossed out”, by contrast, was mostly used to describe things that would elicit core-disgust such as blood, vomit, and feces. Clearly, socio-moral violations do evoke both disgust and anger to some extent. Nabi’s (2002) experiment did not, however, directly account for the relative amount of disgust and anger that can be evoked by socio-moral acts versus immoral acts that involve one’s body.

This comparison between socio-moral acts and acts that involve one’s body was explicitly examined by Gutierrez, et al. (2012) who sought to examine the extent to which disgust measures, separate from anger measures, would vary according to the type of moral violations being described. A further innovation of Gutierrez and colleagues (2012) is that they attempted to resolve the semantic uncertainty between anger and disgust by measuring the two emotions with both standard word-item measures as well as facial expression agreement measures. By doing so, they expected that participants would be able to more effectively report their emotional reactions to the different manipulations, or at least inconsistencies

in participant endorsements of facial versus word-item measures would become apparent in order to inform on the role of anger and disgust in different moral contexts.

Their results indicated that when socio-moral harm was caused, but the body was not used in an immoral way, participants' use of disgust language was largely predicted by their use of anger words, and to a lesser degree by their endorsement of disgust faces. In contrast, when participants read vignettes in which no harm was caused, but bodily moral norms were violated, the use of disgust language was equally predicted by both anger words and disgust faces, thus suggesting that while participants may use disgust language across both types of acts, their endorsement of disgust faces was more present in the bodily moral condition. By demonstrating the extent to which these two types of acts were associated with the facial expression of disgust, the researchers were not confined by the semantic similarities of anger and disgust. As a result, they were able to successfully demonstrate that when an act involves one's body in an immoral way, the disgust that people feel is not just an anger synonym, it is a distinct emotion. As indicated by Gutierrez et al.'s (2012) experiment, the key element of this distinction lies in rules that govern norms regarding the use of our bodies, which leads to the final position on moral disgust that was proposed by Russell and Giner-Sorolla (2013), the bodily norm position.

The three aforementioned positions on moral disgust all fail, to a certain extent, to clearly demonstrate the role that disgust plays in a moral context. The general morality position argues that disgust does play a role in morality, but it lacks specificity and does not consider the well documented

co-activation of anger and disgust (e.g. Gutierrez et al., 2012; Shaver et al., 1987). The metaphorical use position argues that this co-activation of anger and disgust is entirely responsible for the role that disgust plays in the moral realm. According to this position, the disgust language that people use when condemning immoral behaviors is nothing more than a by-product of anger, which supporters of this position argue to be the most relevant other-condemning moral emotion (Nabi, 2002). Thirdly, the purity position comes close to hitting the mark in that it argues that moral disgust is triggered in response to acts that are seen as contaminating one's soul, body, or mind (e.g. Rozin et al., 1999). This position does, however, fall short of offering a satisfactory definition of moral disgust because while the link between bodily impurity and disgust has been thoroughly demonstrated (Russell & Giner-Sorolla, 2013), it has not been satisfactorily shown if the non-bodily aspects of purity (i.e. spiritual and mental purity) can evoke disgust in and of themselves.

In an attempt to resolve this inconsistency in the purity position, and to more generally establish a distinction between acts that evoke both disgust and anger, Russell & Giner-Sorolla (2013) proposed moral disgust is, most specifically, evoked from acts that violate bodily norms. Moreover, violations of bodily norms should allow for the most notable distinction between anger and disgust to present themselves. For when these emotions co-occur, as they would when an act is both harmful and a violation of bodily norms (i.e. bestiality or sexual abuse), their levels should vary independently of one another because the disgust would be primarily evoked from bodily-norm

violation rather than simply being used as a metaphor to express anger at the co-occurring violation of another's rights.

The malleability of anger and the rigidity of disgust

The ability of these emotions to co-occur, yet vary independently, has been demonstrated in a number of experiments (Gutierrez et al., 2012; Russell & Giner-Sorolla, 2011b, 2011c) and it highlights a critical difference between the emotions of anger and moral disgust that gives them further distinction: their respective flexibilities. At first glance, disgust may seem to be a flexible emotion considering that it has a large array of elicitors that ranges from all types of foul stimuli to all sorts of immoral behaviors. In truth, however, it seems that the elicitation of disgust is instead inflexible, even irrational, and not easily swayed by contextual information.

An earlier demonstration of the inflexibility of core disgust was shown by the previously mentioned experiment of Rozin et al., (1986). Rozin and colleagues demonstrated how participants were disgusted by juice that had come into contact with a disinfected cockroach but that the disgust was spread more generally, and clearly irrationally, to a fresh serving of the juice that was poured into a new glass. Similar irrationality has also been shown in undergraduate participants that were asked to evaluate products in a shopping cart. The researchers placed objects that were of a disgusting nature (although entirely sanitary) in close proximity to a variety of target products that ranged from cookies to notebooks. It was found that the presence of the disgust-eliciting products lowered participant evaluations of the other products in the shopping cart. What is more, direct contact of the

two products enhanced participants' negative ratings, but negative evaluations occurred even in the absence of direct contact (Morales & Fitzsimons, 2007). These two examples demonstrate how disgust reactions are inflexible and even categorical. Once something has been branded as disgusting, or has been associated with something disgusting, it is very difficult to change one's opinion of it. This is not, however, the case for anger, which has been shown to be a more flexible emotion that can be mitigated by contextual information and subjected to reappraisal (Russell & Giner-Sorolla 2011c).

For instance, imagine a young boy who decides to prank his father by placing fake vomit on the floor for his dad to find in the morning. When the father wakes up and finds the vomit on the floor of the kitchen, he may be physically disgusted by the object and angry at the dog, who he assumed made such a terrible mess. When he realizes that the vomit is made out of rubber, however, the anger would dissipate and he may even start to laugh. The disgust, on the other hand, would most likely persist to some degree and he would feel a twinge in his stomach when he picked up the fake vomit.

Considerations such as these have led to the conclusion that disgust is an object-based emotion that is acquired through associative learning rather than through appraisals (Giner-Sorolla, 2012; Russell & Giner-Sorolla, 2013). If disgust followed the typical pattern set by appraisal theories of emotion then the disgust that is felt towards fake vomit should disappear when we realize that it is not the foul contaminant that it was initially thought to be. Similarly, Rozin et al.'s (1989) participants should not have rejected a new serving of juice in a clean and cockroach-free glass. It seems, however,

that the disgust evoked from these objects is deeply ingrained. It is not entirely clear when humans develop their disgust responses, but people seems to have a concrete dislike of certain tastes from infancy (Cowart, 1981), and by roughly five years of age people have more fully developed their association between disgust and its classic core elicitors such as blood and bodily waste (Stevenson, Oaten, Case, Repacholi, & Wagland, 2010). Indeed, it seems that people learn to associate certain objects with disgust and once these cognitions have been established, these items are tainted in our eyes and seeing them will unequivocally generate disgust even if we are aware of it being an irrational reaction to a sanitary object.

The inflexibility of disgust does not, however, apply only to core disgust for its stubborn nature has also revealed itself in the moral realm. Most notable, Russell and Giner-Sorolla (2011c) explored the extent to which anger versus disgust would be responsible for changes in participants' condemnation towards various immoral acts. Participants were randomly presented with vignettes that described a man committing moral violations that were intended to arouse anger or disgust. For instance, a harm code violation described a man kicking a dog whereas a bodily-norm violation described a man eating a dead dog. Participants reported the extent to which the behaviors were morally wrong, the amount of anger and disgust that they felt towards the vignettes, and critically, they were asked to consider, and list, various circumstances that would change their evaluations of the described act. Lastly, participants were given an opportunity to reevaluate the behavior in light of the changes that they envisaged. The results indicated that anger, but not disgust, responded to participants' reevaluations of the scenarios and

what is more, any change in participants' moral condemnation was predicted by the change in anger but not by a change in disgust.

In this experiment, Russel & Giner-Sorolla (2011c) gave participants the liberty of freely creating any mitigating circumstance that their imagination could muster. This personalization allowed anger ratings to drop, but the disgust they felt towards the bodily moral violations persisted. This emphasizes how the steadfast and inflexible nature of disgust does not only apply to physical disgust, but to moral disgust as well. It should, however, be noted that the varying flexibilities of these emotions have only ever been examined in the context of real life. In order to push the boundaries of their flexibilities, the research that is presented in this thesis will explore the extent to which the malleability of these emotions crosses the line from reality to fiction. When crossing the line from reality to fiction, anger should dissipate seeing as it is easily mitigated by situational factors. Disgust, by comparison, should remain relatively more stable between real and fictional contexts and this should be especially true when the act in question violates bodily moral norms.

It has been discussed how anger and disgust differ in their flexibilities, but the notion that levels of anger and disgust may vary across fictional contexts is more specifically drawn from the extent to which these emotions are impacted by contextual information. A prime example of this was demonstrated by Goldberg, et al., (1999) who showed how the contexts that surround an unfair act of violence can have an effect on subsequent levels of condemnation. For example, the researchers' experiment presented participants with a video that showed a blatantly unfair act of harm -- a man

beating up a helpless teenager – that successfully elicited anger from those who viewed it. This video clip was shown to all experimental groups, but between the groups, the experimenters manipulated the extent to which the violent offender got his comeuppance. Either the perpetrator was punished, avoided punishment, or there was no information regarding what happened to the man. Afterwards, in an “unrelated experiment” participants were presented with vignettes that described negligent and reckless behavior that caused unintentional harm. The authors found that the distribution of justice in the initial scenarios mitigated the extent to which the morally ambiguous acts in the second half of the experiment evoked anger and were subjected to moral condemnation. In a similar vein, when the initial acts went unpunished, subsequent levels of anger and condemnation were increased. These findings inform on the contextual nature of anger in two ways. Firstly, they emphasize the important role that justice appraisals play in the elicitation of anger. More generally, however, they demonstrate how anger towards, and the condemnation of, harmful acts is fickle, subject to change, and modified by the context in which it occurs. This later point is a critical element of the general hypothesis of the fictive pass asymmetry hypothesis.

Another facet of anger that highlights its contextual nature is the fact that the social context that surrounds a harmful transgression can influence the amount of anger that is elicited by the transgression. This was demonstrated by Fischer and Roseman (2007, Experiment 3) who instructed participants to imagine either a close friend or a stranger acting harmfully and irrationally towards an innocent stranger. When the aggressor was imagined as a friend, relative to a stranger, participants reported

experiencing high levels of anger as well as its approach oriented action tendencies in an attempt to modify the friend's behavior.

These contextual elements of anger do not always need to be in regards to anger felt towards other people. Granted, as an other-condemning emotion (Rozin et al., 1999), anger is usually thought of in a dyadic or group-based setting. Just as often, however, one can feel anger towards the self (Ellsworth & Tong, 2006) and when it is self-directed, anger can similarly be mitigated or alleviated by contextual factors (Hirsch, Webb, & Jeglic, 2012).

Contrary to the contextually sensitive nature of anger, disgust's inflexible and object-based nature makes it so it is not mitigated by contextual factors. One experiment that demonstrates this phenomenon was conducted by Olatunji, Forsyth, & Cherian (2007) in which the researchers conditioned participants to associate words with images that were either neutral (i.e. office supplies) or graphic and disgusting (i.e. bodily mutilation). The graphic images evoked various negative emotions such as panic, fear, and disgust. Over time, however, the amount of fear and panic that participants felt in response to seeing the trigger words decreased. Levels of disgust, on the other hand, did not dissipate quite so easily.

The previous example is more relevant to concerns of core disgust than moral disgust, but it has been demonstrated how bodily moral disgust is also resilient to contextual information. Consider an experiment by Gutierrez and Giner-Sorolla (2007) in which the researchers describe a scientist who invites friends over for dinner. In one condition, meant to evoke anger, the scientist secretly served her dinner guests a harmless memory-enhancing

powder. As expected, this violation of trust resulted in participants feeling mostly anger towards the scientist. In contrast, disgust was the most prominent reaction to another scenario in which the scientist instead served her dinner guests cloned steak, created from cells that she had extracted from her own body. A follow-up experiment expanded upon this story setting by adding further contextual elements such intentionality to the story (Russell & Giner-Sorolla, 2011a). Intentionality was manipulated by describing the powder or steak as being served as the result of a blameless mix-up or on purpose. In line with the non-contextual aspects of disgust, there was no effect of intentionality on participants' levels of disgust when the cloned steak was the dish being served. Participants were disgusted by the harmless cannibalism, regardless of whether or not the steak was served intentionally or by accident.

From an evolutionary point of view, having levels of core disgust be unaffected by contextual factors is an adaptive function that helps us remain healthy and disease free. After all, it is better to register a sanitary or non-pathogenic item as disgusting and be wrong, than to run the risk of contamination (Oaten, et al., 2009). This "better safe than sorry" heuristic is useful in some regards, but it can backfire and result in the stigmatization and marginalization of people who, by no fault of their own, have abnormal morphologies such as amputations, or birthmarks (Park, Faulkner and Schaller 2003; Park, van Leeuwan, Chochorelou, 2013). It has, however, been suggested that these types of reactions may be inadvertent and that evaluations of these stigmatized populations may become relatively more

positive when given enough time for people's reason to overrule their gut reaction (Pryor, Reeder, Yeadon, Hesson-McInnis, 2004).

Conclusion

In closing, it is important to reiterate the fact that the emotions of anger and disgust share linguistic similarities, and often times these two emotions are uttered in the same breath when one is expressing one's outrage. Sure enough, the linguistic similarities between these emotions have been empirically demonstrated (Shaver et al., 1987) and some have even put forth the notion that when these terms are expressed towards moral transgressions, disgust is merely a synonym for anger (Nabi, 2002). More contemporary research on this topic has attempted to dispel the supposed similarities between anger and moral disgust by arguing that the difference between them most clearly presents itself when the moral transgression in question involves an abnormal and immoral use of one's body (Russell & Giner-Sorolla, 2013). Furthermore, critical differences between these two emotions are revealed when one considers the extent to which they are flexible and able to be mitigated by the contexts that surround them. Anger, more so than disgust, is a flexible emotion that is associated with harmful behavior (Russell & Giner-Sorolla, 2011a; 2011b), whereas disgust is a less flexible emotion, resilient to contextual information, and typically evoked from impure acts; specifically those that involve an abnormal use of the body (Piazza, et al., 2013; Russell & Giner-Sorolla, 2013). Considering the extent to which these emotions can be effected by contextual information, the forthcoming experiments will demonstrate how anger, when it is directed

towards fictional relative to real behavior, should drop more than disgust does. In other words, when one engages with fiction that encourages its user to imagine harmful acts and impure acts, people should not feel very much anger towards this person, especially compared to the amount of anger that would be evoked from the act being committed in the context of real life. The diminishing anger should, therefore, leave disgust as the most prominent emotional reaction to the fictional behaviors, especially when it displays an act that violates bodily moral norms.

CHAPTER 4

AN INITIAL INVESTIGATION OF EVALUATIVE ASYMMETRIES ACROSS CONTEXTS

The following chapter reports three experiments that show the systematic development of methodology that established initial evidence of an evaluative discrepancy between different types of morally relevant acts across different contexts (i.e. the fictive pass asymmetry). The first experiment was the first investigation of the fictive pass asymmetry hypothesis. As such, it did not specifically compare harm and purity code violations across contexts. Instead, Experiment 1 took a more general approach to the asymmetry by instead examining emotional reactions and moral judgments towards real versus fictional acts that were of a more general violent versus sexual nature.

Of course, the specific predictions of the fictive pass asymmetry hypothesis are based on comparisons of strictly harmful (harm to another person) versus impure (a violation of bodily moral norms) acts. This initial experiment, however, opted to conduct a more general exploration of the asymmetry by comparing acts that were either violent or sexual in nature, without focusing too directly on the specific moral norms that were being violated. This approach may have reduced experimental control, but it allowed for a better initial understanding of how people morally evaluate real versus fictional immoralities in a general sense.

As discussed in the previous chapters, evidence suggests that the reactions to these behaviors should differ depending on the nature of the

described act and the context in which it occurs. Violent acts should primarily evoke harm concerns whereas sexual acts should primarily evoke purity concerns. As such, the violent acts, more so than the sexual acts, should be mitigated by the context in which they occur (Gutierrez & Giner-Sorolla, 2007). Similarly, anger responds to contextual factors to a greater extent than disgust does (Russell & Giner-Sorolla, 2011b). It was therefore predicted that violent acts would be relatively more acceptable in fictional contexts than in real life. Sexual acts, on the other hand, should show a less substantial reality to fiction drop in moral condemnation. Similarly, anger should show fictive pass asymmetry because of its more flexible nature whereas disgust should remain more constant across the different contexts.

Following this initial exploration of the fictive pass asymmetry, two more experiments will be presented that do away with the general categorization of violence versus sex and instead use the more theoretically valid classifications of harm versus purity code violations. By doing so, these experiments will demonstrate that fictive pass effects remain, and are even more pronounced, when the vignettes that describe the immoral behaviors draw a clear distinction between the moral codes of harm and purity. Together, these experiments will lend initial support to the fictive pass asymmetry hypothesis by demonstrating the role that contextual cues play in evoking varying degrees of emotion and moral condemnation.

Participant recruitment

In the experiments that are presented in this chapter, and all subsequent chapters, participants were recruited online from Amazon's Mechanical Turk Service (https://mturk.com). Since this service was used to

collect data for all of the experiments presented in this thesis, the reliability and validity of Mechanical Turk (MTurk) as a means of participant recruitment will be briefly discussed before moving forward with presenting the first series of experiments.

MTurk is a crowdsourcing platform that allows individuals or organizations to recruit large numbers of people to complete tasks online. MTurk was not originally created as a means for scientists to gather data, but rather as a way for companies to outsource low skill and low effort tasks like date entry, transcription, or image identification that cannot be effectively done by computers. Eventually, researchers identified MTurk's potential as a fast and cost effective tool of participant recruitment and by about 2010, it started seeing regular use in the social sciences (Paolacci, Chandler, & Ipeirotis, 2010). One may be duly skeptical towards a platform that can potentially recruit hundreds of participants in a few hours. Yet, replication studies have been conducted on MTurk and the results suggest that MTurk data is just as reliable as data from more traditional sources, such as an undergraduate population (Paolacci et al., 2010). Moreover, MTurk's large user base is more representative of the USA population than an undergraduate participant pool is (Buhrmester, Kwang & Gosling, 2011), it allows data to be collected faster, and its cost effectiveness means that one can recruit large samples for high powered experiments, all without compromising data quality. There are potential downsides and consequences to using this service, and the extent to which these effected the data of these experiments will be considered in the general discussion. Nonetheless, MTurk is a reliable means of data collection and for the

previously discussed reasons, the experiments within this thesis will rely upon it as a means of participant recruitment.

Experiment 1

Experiment 1 compared anger, disgust, and moral judgments of act and character towards people who commit moral transgressions in real life versus those who consumed the same act through a form of media (watched in a film or performed in a video game). The emotions of anger and disgust were examined because their relative flexibilities (Russell & Giner-Sorolla, 2011b) should allow them, in conjunction with the moral judgments, to highlight the effects of the fictive pass asymmetry. The primary aim of this experiment was to identify significant Violation Type (violent vs sexual) x Context (real vs fiction) interactions that would indicate a larger evaluative discrepancy between reality and fiction for violent acts than for sexual acts.

Method

Participants

The data of 431 USA residents (296 Male; M_{age} = 32.17: SD_{age} = 10.23) were collected from Amazon's Mechanical Turk service. A total of 515 participants were recruited, but 84 (16%) were excluded because they failed an attention checking question that asked them to report if the act they read about was committed in real life, watched in a film, or performed in a video game.

Design

This experiment employed a 2 (type of offense: violent vs sexual) x 3 (context of violation: reality; watched in a film; performed in a video game) between-subjects design. Participants were presented with a single scenario that described an act of violence or an abnormal sexual act that was further randomized to be presented as occurring in real-life, as watched in a film, or as performed in a video game. There were no specific hypotheses regarding the differences between the two fictional contexts, so all of the fictional contexts were collapsed into a single level of fiction. Consequently, the data was analyzed as 2 (type of offense: violent vs sexual) x 2 (context of violation: reality vs fiction) experiment.

Materials

In designing this experiment, multiple vignettes were created to portray both the violent and sexual acts. The violent vignettes described one person committing direct harm to another and the sexual vignettes described bizarre sexual acts.

These vignettes can be found in their entirety in the supplementary materials A but examples from each subcategory can be found below. The below examples are all set in the context of real-life, but one can readily understand how the wording was manipulated to present these acts as something that the main character watched in a film or performed in a video game.

Violent Vignettes

Peter likes to attack elderly people with a device that will cause their pacemakers to malfunction.

Marshal intentionally ran over a pedestrian at a street crossing. He heard a loud crunch and squishing sounds as the man went under the tires.

Sexual Vignettes

Tom is 57 and only finds 16-year old girls attractive. He won't sleep with anyone under 16 as he knows it's illegal but he will find girls as young as possible to have sexual relations with.

Clark is a frotteur. This means that he goes into crowded public places like subway stations and buses and attempts to rub his erection against unsuspecting victims.

To measure condemnation, self-report measures assessed the extent to which the acts described in the vignettes evoked negative emotions and were morally condemnable. To address the typically strong correlation between anger and disgust, these emotions were measured with both word item measures and facial expression agreement measures (as recommended by Russell & Giner-Sorolla, 2011b). Word items asked participants how much the scenario evoked the target emotion (i.e. *disgust* and *anger*) and two synonyms of each emotion (*furious, outraged*; *sickened, revolted*) on a scale from 1 (*not at all*) to 7 (*entirely*). Facial expression agreement measures presented participants with two separate sets, each of three pictures, that displayed female faces that were expressing the target emotion at full intensity (Beaupré, Cheung, & Hess, 2000). Each set of faces was labelled with their respective emotion and participants reported how

much the scenario made them feel like the displayed emotion on a scale from 1 (*not at all*) to 7 (*entirely*) (see Figure 1). As discussed in chapter 3, the emotions of anger and disgust share many similarities, especially in a moral context. This method of measurement should allow participants to differentiate between them and reduce the covariance that tends to arise from the semantic similarities between moral disgust and anger (Russell & Giner-Sorolla, 2011b; Shaver et al., 1987).

Figure 1. Facial expression agreement measures of anger and disgust. From Beaupré, et al., (2000).

To measure moral judgments, two scales assess two critical facets of moral condemnation: act-based and character-based judgments (Chakroff & Young, 2015). Three items measured the moral wrongness of the main agent's actions (*How moral is [Agent's] behavior?*; *How wrong is [Agent's] behavior?*; *How right is [Agent's] behavior?*; α = .83) and three items focused on the moral status of the agent's character (*Do you think that Agent is mainly a good or mainly a bad person?*; *How trustworthy is [Agent]?*; *Do you think that [Agent] has good moral standards?*; α = .84). All items were coded so that higher numbers equaled more condemnation. It was acknowledged in

chapter 2 that simple measures of character such as these may undermine its multi-faceted nature (Goodwin et al., 2014). This issue will be addressed in chapter 6's Experiment 6.

Table 1.

Correlations, means, and standard deviations of all dependent variables, Experiment 1. All correlations are statistically significant at p < .001.

	1.	2.	3.	4.
1. Anger				
2. Disgust	0.86			
3. Act-based Judgments	0.63	0.71		
4. Character-based Judgments	0.61	0.64	0.82	
Mean	3.78	4.36	5.12	5.02
SD	1.98	1.96	1.58	1.32

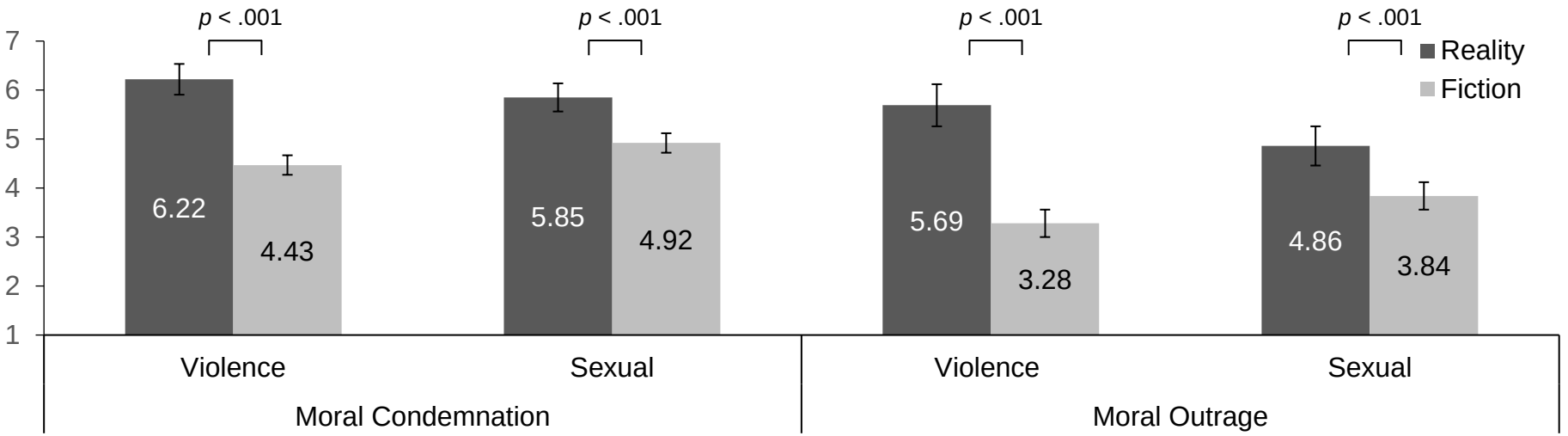
Results

The anger and disgust scales (r = .86, p < .001) and the moral wrongness and moral character scales (r = .82, p < .001) were both highly correlated. To address this, two composite items were created. The emotion items formed the new *moral outrage* item and the moral judgment items became the new *moral condemnation* item. These new items had correlation that was still very strong, but it was somewhat weaker (r = 0.70, p < .001) than the original composites and thus these new composite items were analyzed separately.

For the main analysis, a series of 2 (type of act: violent vs sexual) x 2 (context of violation: reality vs fiction) univariate ANOVAs were conducted to

test for fictive pass asymmetry effects across the dependent variables. These analyses were primarily seeking significant Type x Context interactions which would indicate that the evaluative drop from reality to fiction was greater for violent acts than it was for sexual acts.

Across the two dependent variables (*moral outrage, moral condemnation*), significant Type x Context interactions supported the fictive pass asymmetry hypothesis by indicating that fictional acts of violence are significantly more acceptable than the same acts that occur in the context of real-life (Figure 2). By contrast, this evaluative discrepancy was relatively smaller when comparing evaluations of real versus fictional sexual acts. In other words, to consume violent fiction is less immoral and evokes less negative emotions than acting violently in real-life. This reality to fiction drop is, however, significantly smaller when the fiction in question was of a sexual nature. That being said, this difference is relative and not absolute. Simple effects for moral condemnation indicate that the reality to fiction drop is significant for both violent acts, $F(1, 427) = 91.50$, $p < .001$, $\eta_p^2 = .18$, and sexual acts, $F(1, 427) = 26.92$, $p < .001$, $\eta_p^2 = .06$. For moral outrage, a similar pattern emerged for violent acts, $F(1, 417) = 85.52$, $p < .001$, $\eta_p^2 = .17$ and sexual acts, $F(1, 427) = 17.28$, $p < .001$, $\eta_p^2 = .04$. An absolute difference between these different types of acts would be the strongest expression of the predicted fictive pass asymmetry effect. Nonetheless, the overall hypotheses are supported in that a relative lenience towards fictional violence, but not sex, has been identified, indicated by the main interaction effects as well as each simple effect's partial eta squared,



Contrasts	Dependent Variable		Type Main Effect	Context Main Effect	Interaction Effect
Type	Context				
Violent vs Sexual	Reality vs Fiction	Moral Condemnation	$F(1, 427) = 0.06$, $p = .81$, $\eta_p^2 = .000$	$F(1, 427) = 110.75$, $p < .001$, $\eta_p^2 = .21$	$F(1, 427) = 11.68$, $p = .001$, $\eta_p^2 = .03$
		Moral Outrage	$F(1, 427) = 0.59$, $p = .44$, $\eta_p^2 = .001$	$F(1, 427) = 91.90$, $p < .001$, $\eta_p^2 = .18$	$F(1, 427) = 15.15$, $p < .001$, $\eta_p^2 = .03$

Figure 2. Moral condemnation (average of moral wrongness and moral character) and moral outrage (average of anger and disgust) across all type of acts and contexts, Experiment 1. Means, 95% confidence intervals, and simple effects are presented in chart. Higher numbers equal greater moral condemnation or affect.

Discussion

The results of Experiment 1 have lent initial support to the fictive pass asymmetry hypothesis by showing a discrepancy in the extent to which acts that are of a violent nature or a sexual nature are condemned across different contexts. More specifically, significant interactions indicated that the evaluative discrepancy between real and fictional acts of violence were significantly larger than the gap between real and fictional sexual acts.

Experiment 1 tested the fictive pass asymmetry hypothesis with manipulations that were generally categorized as behaviors that were either violent or sexual in nature. Due to this broad classification, a number of these acts varied in the extent to which they were grotesque, consensual, bizarre, and intentional. On the one hand, this experiment demonstrated fictive pass asymmetry effects across a diverse number of facets that are relevant to moral condemnation but on the other hand, this variety resulted in a lack experimental control.

For instance, one violent act described a doctor who preferred to euthanize his patients over giving them proper treatment while another described a man who shot another person with a shotgun. Both these acts can be broadly categorized as violent and while this highlights the breadth of behaviors that may demonstrate fictive pass asymmetry effects, certain details of these manipulations make them differ from one another in ways that are potentially problematic. The act involving the shotgun, for instance, elicits concerns of core disgust by evoking imagery of blood and gore, whereas the vignette involving the doctor does not contain any elicitors of

physical disgust, but describes a typically high status member of society breaking the trust that patients give them. In a similar vein, one of the sexual vignettes described a frotteur who sexually assaults people without consent, while another sexual vignette described an older man who sought sexual relationship with barely legal teenage girls, but it was explicitly stated that the relationship between them was consensual.

In effect, the content of the violent and the sexual vignettes blurred the lines of their most closely related moral codes of harm and purity (as discussed in chapter 2), and this may explain why there were such strong correlations between the dependent variables. Nonetheless, fictive pass asymmetry effects emerged as expected which may be a testament to the validity of the hypothesis. Nonetheless, when moving forward it is important to refine the manipulations by examining the fictive pass asymmetry in line within the confines of the established literature on the moral domains (i.e. Graham, et al., 2011; Rozin, et al., 1999).

Experiment 2, therefore, utilized a new batch of vignettes. Rather than vaguely focusing on violence versus sex as manipulations, the content of Experiment 2's vignettes abided by the established moral codes and strove for greater parallelism across these moral domains.

Experiment 2

Experiment 1 lent initial support to the fictive pass asymmetry hypothesis by showing that fictional acts of violence, compared to fictional sex, were relatively more acceptable than their real life counterparts. There

are, however, a number of points that need to be addressed regarding the nature of the stimuli that were used in the first experiment.

Firstly, the intentional use of the broad categories of violence and sex did not fully acknowledge, and even went so far as to confound, harm and purity code violations as they are defined by the leading moral psychology theorists (Graham, et al., 2012; Russell & Giner-Sorolla, 2013). This is an issue that needs to be addressed because even though the wide breadth of Experiment 1 had its advantages, the fictive pass asymmetry hypothesis makes specific predictions about the nature of acts that cause harm to other individuals versus impure acts that violate bodily moral norms. In Experiment 1, violence was not always free of impurities, and sexual acts were not always free of harm. For instance, consent, or the lack thereof, is a common feature of prototypical harm, so it makes sense that it was present in many of the violent vignettes. A lack of consent was, however, also present in some of the sexual vignettes which allows for the possibility that participants' reactions to them were confounded by the elements of harm that they contained. Similarly, some of the violent vignettes contained gory imagery that would elicit feelings of physical disgust, thus confounding the harm and purity domains. As a result, these two moral domains were conflated, perhaps even more than they were individually examined.

Furthermore, in addition to confounding these two moral domains, the violent and the sexual vignettes lacked parallelism. As such, one can question the extent to which the fictive pass asymmetry effects were due to the specific behaviors that described the violent and sexual acts, rather than an actual discrepancy in how they are evaluated across contexts.

To address these issues, the manipulations of Experiment 2 aimed to depart from the vague categories of violent and sexual acts and instead focus more directly on acts that violate the harm norms (Graham et al., 2012) and bodily moral norms (Russell & Giner-Sorolla, 2013). For the sake of clarity, and to be in line with most of the literature on moral domains, bodily moral norm violations will be referred to as purity code violations.

Method

Participants

The data of 360 United States residents were collected from Amazon's Mechanical Turk service. 89 (24.7%) participants were excluded because they failed an attention checking question that asked them to report if the act they read about was something that was done in real life, watched in a film, or performed in a video game. Participants who had taken part in Experiment 1 were not able to participate because of similar hypotheses and methodology. Ultimately, the data of 271 United States residents (179 male; M_{age} = 31.56; SD_{age} = 9.54), collected through Mechanical Turk, was included in the final dataset.

Design

Experiment 2 employed a 2 (code violation: harm vs purity) x 3 (context of violation: reality vs watched in a film vs performed in a videogame) between-subjects design. As in Experiment 1, there were no specific hypotheses regarding the differences between the fictional contexts, so for analysis the two media products were collapsed into a single level of fiction. Participants were randomly presented with a single vignette that

described an act that violated moral norms of harm or moral norms of purity. Further randomization described the act as being committed in real-life or as being enjoyed in one of the two fictional contexts (film or video game).

Materials

Experiment 2 sought to resolve the issues of Experiment 1 by making clearer partitions between the two relevant moral domains and by creating more parallelism between the vignettes that described each type of act.

To maintain parallelism, three general story settings were written and then critical elements of each story were modified in order to present the act as one that violated moral norms of harm or moral norms of purity. Harm vignettes described a main character acting in a way that caused direct harm to another individual and these acts did not contain any elements of physical disgust or unnatural uses of one's body. Purity code violations described a main character committing an act that violated bodily moral norms and these acts were entirely consensual and free of interpersonal harm. The harm and purity vignettes are displayed below, presented as they appeared in the context of reality. Below these, there are two additional examples, one from the film condition and one from the video game condition. Between these examples, one can see how any code by context combination would have been portrayed, but the manipulations can be found in their entirety in supplementary materials A.

Harm Vignettes

Robert is a university student who owns a piercing gun. He goes to parties and secretly gives small and tasteful ear piercings to people who are passed out drunk.

Robert likes to humiliate his girlfriend by saying cruel and abusive things in public.

Robert, a 19-year old, works at a retirement community and he deceives and threatens an old woman to inherit her legacy.

Purity Code Vignettes

Robert is a university student who owns a piercing gun. He goes to parties and enjoys giving genital piercings to anyone who wants one.

Robert likes to humiliate his girlfriend with degrading sex acts that she consents to.

Robert, a 19-year old, works at a retirement community and has a consensual sexual relationship with a 72-year old woman.

Watched in a Film Example

Robert, a 19-year old, watched a film about a 19-year old man who works at a retirement community and has a consensual sexual relationship with a 72-year old woman. Robert enjoys watching this film.

Performed in a Video Game Example

Robert, a 19-year old, plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Robert can control his character to do in this virtual environment.

In the video game, Robert's character works at a retirement community. Robert controls his character to have a consensual sexual relationship with a 72-year old woman. Robert enjoys playing this game.

The emotion measures were unchanged from Experiment 1. On scales that ranged from 1 (*not at all*) to 7 (*entirely*), participants reported the amount of anger and disgust that they felt via word-item measures – angry, furious, outraged, disgusted, sickened, revolted – as well as facial expression agreement measures that showed black and white photos, three for each emotion, of women expressing anger and disgust at full intensity (Beaupré, et al., 2000). The anger scale ($\alpha = .91$) and the disgust scale ($\alpha = .91$) both had strong reliabilities. They also had an unusually high correlation of $r = .87$, $p < .00$ that was equal in strength to the anger and disgust correlation of Experiment 1. This happened in spite of efforts to make a clearer partition between harm and purity code violations. A composite item of *moral outrage* was formed from the anger and disgust scales.

The act and character-based moral judgment items were differentiated with one scale that measured the agent's moral character (*Is [Agent] "rotten inside"?*; *Is [Agent] 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 scale that measured the morality of the act that the main agent 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 scales, the moral judgment scales also had an extremely high correlation of r = .85, p <. 001 just as they did in Experiment 1. As such, another composite item was created to assess *moral condemnation.* The correlation between the *moral wrongness* and the *moral outrage* variables was, however, also extremely high at r = .83, p < .001. Considering that the strong correlation between these two composite items was no lower than the correlations of the original scales, all of the dependent measures were collapsed into a single item that assessed general *negativity* rather than specific emotions or moral judgments.

The resulting measure of *negativity* was analyzed using an ANOVA that crossed the type of violation with the context in which it was committed in a 2 (moral code: harm vs purity) x 2 (context of violation: reality vs fiction).

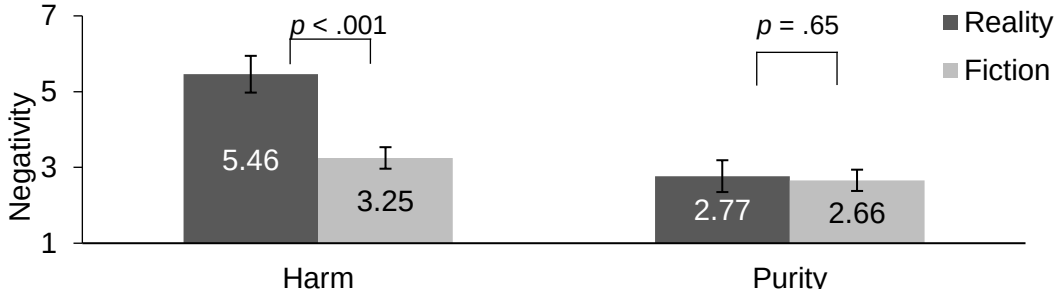
Results

Main effects indicated that real acts and harmful acts were most negatively evaluated. In support of the fictive pass asymmetry hypothesis, a significant Code x Context interaction revealed that fictional harm was perceived less negatively that real harm, but that fictional and real impurities did not differ in their perceived negativity (See Figure 3). Simple effects emphasized this further by indicating that evaluations towards harm were significantly less negative in fiction than in reality, $F(1, 267) = 60.92$, $p < .001$, $\eta_p^2 = .18$ but

there was no significant difference in negativity towards real and fictional impurities, $F(1, 267) = 0.20$, $p = .65$.

Discussion

Although Experiment 2 did not offer as clear of a distinction between moral judgments and moral emotions, the results once again lent support to the fictive pass asymmetry hypothesis by showing that acts of real life harm are evaluated significantly more negatively than their fictional counterparts. In other words, fictional harm is given a pass in terms of moral condemnation. In contrast to this, the evaluations of purity code violations stayed about the same across both contexts. Moreover, Experiment 2 generalized these effects over a different set of moral violations that had a greater emphasis on parallelism and accurately representing the moral codes of harm and purity.



Contrasts		Code Main Effect	Context Main Effect	Interaction Effect
Harm vs Purity	Reality vs Fiction	$F(1, 265) = 55.50$, $p < .001$, $\eta_p^2 = .17$	$F(1, 265) = 19.50$, $p < .001$, $\eta_p^2 = .13$	$F(1, 265) = 16.32$, $p < .001$, $\eta_p^2 = .11$

Figure 3. Negativity across all codes and contexts, Experiment 2. Means, 95% confidence intervals, and simple effects are presented in the chart. Higher numbers equal more negativity.

Experiment 3

The manipulations of Experiment 2 generalized from Experiment 1 to encompass different immoral behaviors and they placed emphasis on describing exclusively acts that strictly violated moral norms of harm or of purity, but not both. With these emphases, however, an initial pretesting of the vignettes was neglected. Pretesting was also overlooked because it was not an issue in Experiment 1. Indeed, looking at Experiment 1's results (Figure 2), demonstrates that real violent acts (M = 6.22) and real sexual acts (M = 5.85) are about as morally condemnable as one another. This may be due to the fact that violent and sexual categories contained elements of both harm and purity code violations, which served to equalize the condemnation that was directed towards them.

In Experiment 2, however, the behaviors that described impure acts were simply not very condemnable. This is most noticeable when comparing the perceived negativity of real harm (M = 5.46) to the perceived negativity of real impurities (M = 2.77) (Figure 3). This discrepancy in real life condemnation raises the possibility that Experiment 2's fictive pass asymmetry effects were caused by a floor effect for the purity condition. In other words, the negativity that was directed towards real-life harm was substantially greater than what was directed towards real-life impurities. This gave evaluations of harm a greater distance to fall in the fiction condition and may have artificially inflated the effects of the asymmetry.

This should not, however, raise doubts about the legitimacy of this experiment. To examine this data in a different way, another analysis was

conducted that excluded all mean responses of the negativity variable that were less than 2.00. Results indicated that the fictive pass asymmetry still presented itself, although with admittedly weaker effects, $F(1, 183) = 9.27$, $p = .003$, $\eta_p^2 = .05$.

This discounts the possibility that a floor effect of the purity condition was entirely responsible for the results of the fictive pass asymmetry, although it cannot be denied that the floor effect did contribute to the strength of the effect. In light of this, it would be useful to test the fictive pass asymmetry with harm and purity code violations that were relatively more similar in real life moral wrongness, or with real life moral wrongness that was at least closer to the scale midpoint, thus allowing a greater range of responses. To address this, Experiment 3 used a set of newly pretested vignettes and also added a new level of fiction *imagination* in order to generalize the effects of the fictive pass asymmetry beyond media products.

Method

Pretest and vignette selection

For the pretest, forty-two vignettes were written that described violations of harm or violations of purity. As in Experiment 2, harm violations did not involve bodily norm violations and impure acts were entirely free of harm. United States residents ($N = 321$) were recruited from Amazon's Mechanical Turk service and were randomly presented with a selection of six vignettes. Each vignette was followed by a forced choice facial expression agreement measure for anger and disgust using the pictures provided by 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*). In order to keep the pretest quick and simple, no demographic information was collected.

The literature on moral condemnation and moral emotions contains a fair amount of research that has drawn specific links between anger and harm code violations and disgust and purity code violations (i.e. Goldberg, et al., 1999, Horberg et al., 2009, Piazza, Russell & Sousa, 2013, Russell & Piazza, 2015). Drawing from this literature, two harm vignettes and two purity vignettes were ultimately selected that elicited the expected emotions (e.g. purity violations that were higher in disgust than anger and vice versa), and that, critically, were similar in moral wrongness. These selection criteria should allow for two improvements over Experiment 2. Firstly, anger and disgust should be more distinct, which will allow them to be examined separately from another. Secondly, ensuring that the acts are morally equivalent in the context of real life should ensure that emergent fictive pass asymmetry effects can be attributed to the manipulation of the context variable, not the baseline wrongness of any given code violation. 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, $F(3,61) = 0.44$, $p = .73$. The selected vignettes are below, described as occurring in the context of reality.

Harm Code Vignettes

A woman named Sam gets into a fight with her neighbor over the removal of a tree that is on both of their properties. Sam decides

to destroy her neighbor's lawn by driving her heavy truck all over it.

Sam shouted at his girlfriend because she did not have enough time to put on make-up before a date.

Purity Code Vignettes

Sam buys a dead chicken from the store. He masturbates with it before cooking it and eating it.

Sam works at an office with a unisex bathroom. He likes to go into the bathroom after it has been used by female coworkers and lick the toilet seat.

Participants

The data of 321 United Sates residents (229 Male; M_{age} = 31.5, SD_{age} = 9.7) were recruited from Amazon's Mechanical Turk service. Owing to similar measures and manipulations, anyone who had participated in Experiments 1 or 2 was not able to participate. Data of two participants (1%) were excluded for failing an attention checking question that was hidden amongst the dependent measures and explicitly asked participants to select a specific scale point if they were paying attention.

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 three 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 interpersonal harm (property destruction; verbal aggression) that contained no elements or suggestions of bodily moral norms violations. The purity vignettes described moral violations involving the body (sex with a dead chicken bought at the grocery store; bizarre bathroom behavior) that were free of harm to other people. 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.

For Experiment 3, the fictive pass asymmetry hypothesis remained the same. It is predicted that violations of harm will reveal a steep reality to fiction drop in moral violations. Purity violations, on the other hand, should remain more stable across context. Moreover, the fictive pass effects of the moral emotions of anger and disgust should mirror the moral codes that they are most closely associated with; that is, anger, more so than disgust, should show fictive pass asymmetry effects.

Results

The four anger items (3 word item measures and a facial expression agreement measure; α = .94) and the four disgust items (3 word item measures and a facial expression agreement measure; α = .94) were reliable scales that were compiled into their respective composites. Although these composite items had a significant positive correlation, it was much lower than it had been in previous experiments (r = 0.42, p < .001) which allowed them to be analyzed separately. This lower correlation was likely due to the pretest that specifically defined harm versus purity violations as their ability to elicit more anger than disgust and vice versa.

The act-based (α = .91) and character-based (α = .94) moral judgment scales both had strong reliability and each scale was transformed into a composite. The composite items had a correlation that was slightly lower than in Experiment 2, but at r = .77 they still shared over half of their variance. While less than ideal, these items were still analyzed individually in order to begin exploring potential act versus character-based explanations for the fictive pass asymmetry effects (see Table 2 for Experiment 3 correlations, means, standard deviations, and Chronbach alphas).

Table 2.

Correlations, means, standard deviation, and Chronbach alphas of all dependent variables, Experiment 3. All correlations are statistically significant at p < .001.

	1.	2.	3.	4.
1. Act-based Judgments				
2. Character-based	.77			

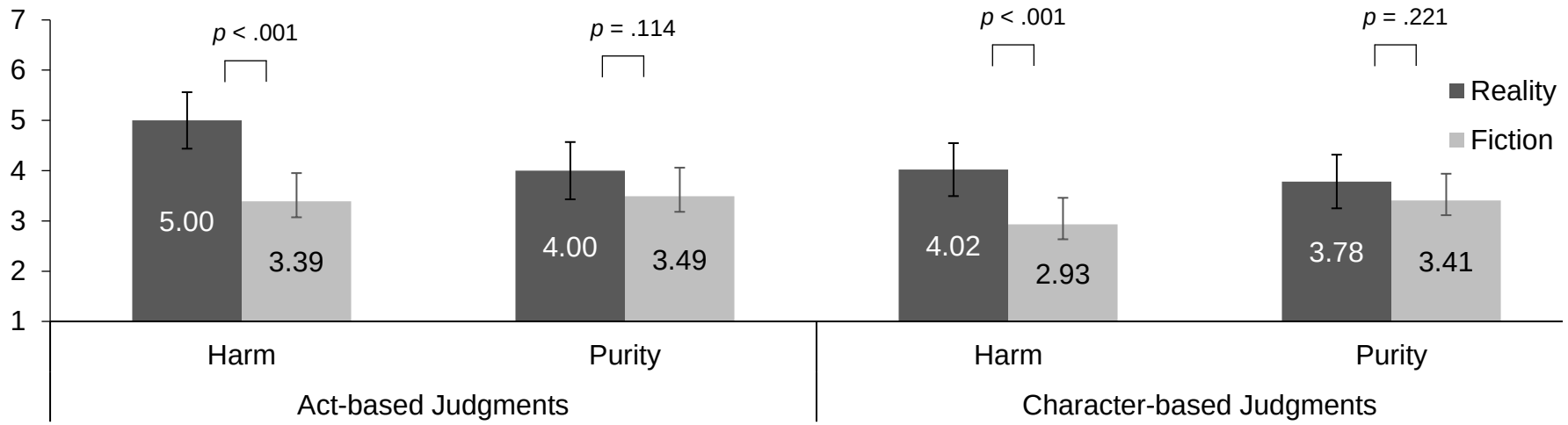
Judgments				
3. Anger	.69	.61		
4. Disgust	.49	.52	.42	
Mean	3.69	3.35	2.93	4.28
SD	1.83	1.69	1.68	1.86
Reliability	α = .91	α = .94	*r* = .94	*r* = .94

In the main analysis, main effects revealed that character-based judgments of real acts did not significantly vary between harm (*M* = 4.02) and purity (*M* = 3.78) code violations. Despite the best efforts of the pretest, however, act-based judgments indicated that real harm violations (*M* = 5.01) were significantly more immoral than real purity code violations (*M* = 4.01), $F(1, 73) = 8.06$, $p = .006$ (Figure 4). In line with pretesting, main effects of the emotion measures showed that harm was more strongly associated with anger and purity was more strongly associated with disgust (Figure 5).

As previously mentioned, an aim of Experiment 3 was to address the floor effects of Experiment 2 by establishing equivalent moral wrongness across the reality condition. Despite successful pretesting, Experiment 3 was not able to achieve equal moral wrongness means between real-life harm and purity violations but, unlike in Experiment 2, the mean wrongness of real impurities was at the scale midpoint (M = 4.00) which allows for a range of responses that is large enough to discount the possibility that floor effects are responsible for any fictive pass asymmetry effects that emerge in Experiment 3.

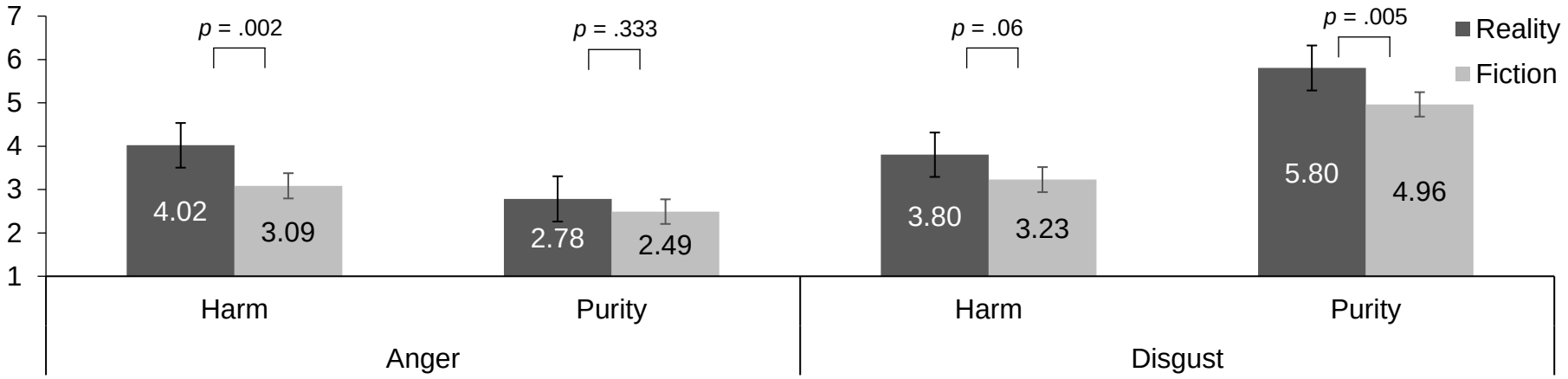
Indeed, a significant Code x Context interaction for act-based judgments did support the fictive pass asymmetry hypothesis. As predicted, fictional harm was more acceptable than its real-life counterpart but this gap between real and fictional contexts was relatively smaller for purity code violations (Figure 4). This effect was further emphasized by simple effects that indicated a significant reality to fiction difference for harm violations, $F(1, 315) = 24.21$, $p < .001$, $\eta_p^2 = .07$, but not for purity violations, $F(1, 315) = 2.51$, $p = .11$, $\eta_p^2 = .008$. Character-based judgments, however, showed weaker fictive pass effects and there was no significant Code x Context interaction. In spite of this, simple effects still indicated a significant reality to fiction drop for harm, $F(1, 315) = 12.51$. $p < .001$, $\eta_p^2 = .04$, but not for purity, $F(1, 315) = 1.50$, $\eta_p^2 = .005$. The lack of an interaction is less than ideal, there is useful information that can be drawn from these results.

Firstly, act-based judgments showed fictive pass asymmetry effects whereas character-based judgments did not. These different effects occurred despite the high correlation of act and character-based judgments ($r = 0.77$), thus reinforcing the decision to analyze them as separate variables.



Contrasts		Dependent Variable	Code Main Effect	Context Main Effect	Interaction Effect
Code	Context				
Harm vs Purity	Reality vs Fiction	Act-based Judgments	$F(1, 315) = 3.81$, $p = .05$, $\eta_p^2 = .012$	$F(1, 315) = 21.10$, $p < .001$, $\eta_p^2 = .063$	$F(1, 315) = 5.52$, $p = .02$, $\eta_p^2 = .017$
		Character-based Judgments	$F(1, 315) = 0.20$, $p = .59$, $\eta_p^2 = .001$	$F(1, 315) = 11.32$, $p = .001$, $\eta_p^2 = .035$	$F(1, 315) = 2.65$, $p = .11$, $\eta_p^2 = .008$

Figure 4. Act-based and character-based judgments across all codes and contexts, Experiment 3. Means, 95% confidence intervals, and simple effects are presented in chart. Higher numbers equal more condemnation.



Contrasts		Dependent Variable	Code Main Effect	Context Main Effect	Interaction Effect
Code	Context				
Harm vs Purity	Reality vs Fiction	Anger	$F(1, 315) = 18.45$, $p < .001$, $\eta_p^2 = .055$	$F(1, 315) = 21.58$, $p < .001$, $\eta_p^2 = .026$	$F(1, 315) = 2.26$, $p = .134$, $\eta_p^2 = .007$
		Disgust	$F(1, 315) = 77.82$, $p < .001$, $\eta_p^2 = .198$	$F(1, 315) = 11.14$, $p < .001$, $\eta_p^2 = .034$	$F(1, 315) = 0.40$, $p = .53$, $\eta_p^2 = .001$

Figure 5. Anger and disgust across all codes and contexts, Experiment 3. Means, 95% confidence intervals, and simple effects are presented in chart. Higher numbers equal more affect.

Secondly, these differences in condemnation suggest that acts of real harm are seen as less indicative of bad character, relative to the wrongness of the act itself. This validates Experiment 3's manipulations by resonating with similar findings that show us how different types of morally relevant behaviors evoke different types of moral condemnation (Chakroff & Young, 2015; Tannenbaum, et al., 2011; Uhlmann & Zhu, 2013).

Moral emotions (Figure 5) showed the expected pattern of main effects in that harm was more strongly associated with anger, and impurities were more strongly associated with disgust. Unlike in the previous two experiments, anger and disgust shared the moderate correlation of $r = .43$ but in spite of them sharing relatively less covariance than before, neither anger nor disgust showed fictive pass asymmetry effects. On the one hand, it was predicted that disgust's trademark inflexibility (Russell & Giner-Sorolla, 2013) would prevent its levels from changing much between real and fictional contexts and in this regard, the hypotheses were supported. Anger, however, being a more flexible emotion (Russell & Giner-Sorolla, 2011a; 2011b), was expected to diminish in fictional contexts to a significantly greater extent than disgust did. These effects were not found in the interactions, but simple effects still indicated meaningful differences between reality and fiction that were congruent with the hypotheses. For anger, harm significantly declined from real to fictional contexts, $F(1, 315) = 9.63$, $p = .002$, $\eta_p^2 = .30$, but anger towards purity violations did not significantly differ between contexts, $F(1, 315) = 0.94$, $p = .33$, $\eta_p^2 = .003$. For disgust, stranger patterns emerged in the simple effects and a marginally significant reality to fiction drop was found for

harm, $F(1, 315) = 3.69$, $p = .06$, $\eta_p^2 = .01$, and a significant reality to fiction drop for purity violations, $F(1, 315) = 7.833$, $p = .005$, $\eta_p^2 = .024$.

Discussion

The results of Experiment 3 lend further support to the fictive pass asymmetry hypothesis by showing that fictional harm is, by and large, more acceptable than fictional purity. Act and character-based judgments were highly correlated ($r = .77$) with each other. This may, however, be a sign that fictive purity violations are disproved of *because* they signal bad character. In fact, this notion was supported by an analysis that excluded real acts and found that the correlation between act and character-based judgments weakens for harm violations ($r = .53$, $p = .001$) but remains about as strong for purity code violations ($r = .70$, $p < .001$).

These findings hint at a unique interplay between fictional impurities and moral character. This notion will be more thoroughly explored in the subsequent chapters.

Conclusion

The three experiments that have been reported in this chapter display the systematic establishment of methodology that ultimately succeeded in revealing the predicted fictive pass asymmetry effects. The first experiment tested the fictive pass asymmetry in a more general sense by comparing acts that described violent versus sexual acts. Despite methodological confounds, the fictive pass asymmetry still presented itself, hinting towards the robustness of these effects. Experiment 2 generalized the fictive pass asymmetry effects to a different set of moral situations and created

manipulations based on established differences between the harm and purity domains rather than the more general categories of violence and sex, and fictive pass asymmetry effects were again found. Lastly, Experiment 3 further refined the methodology with pretested vignettes that best exemplified harm versus purity concerns. Moreover, Experiment 3 broadened its definition of fiction by going beyond media products and including imagination as a level of fiction. Together, these three experiments have provided initial evidence for fictive pass asymmetry effects on moral condemnation and moral emotions but the mechanics that drive these effects are yet to be identified.

CHAPTER 5

CHARACTER-BASED EXPLANATIONS OF THE FICTIVE PASS ASYMMETRY

The goal of the experiments in chapter 5 is to delve deeper into the effects of the fictive pass asymmetry. The two experiments that are reported in this chapter will investigate a number of unexplored, and theoretically fruitful, facets of moral condemnation that should offer a more complete explanation of the fictive pass asymmetry effects.

Experiment 4

Experiments 1 through 3 made consistent methodological improvement and refinements in order to demonstrate the effects of the fictive pass asymmetry. Across these three experiments there has been, more often than not, reliable evidence in support of the hypothesis. Fictional harm has been more morally acceptable than real harm, but this evaluative discrepancy between reality and fiction has been significantly smaller for impure acts that violate bodily norms.

Although fictive pass asymmetry effects have been demonstrated, the exact nature of this condemnation is a bit unclear considering that the moral judgments of act and character have been highly correlated across the three previous experiments. As suggested at the end of the last chapter, this covariance may be an indication that fictional impurities are condemned because they show poor moral character. This conclusion is, however, a little tenuous, and it would be more theoretically sound to compare character

judgments with other facets of moral condemnation that are less strongly affected by the fictive pass asymmetry.

One such facet of moral condemnation that should present itself as more unique from character-based judgments is participants' desire to punish the main agent in the vignettes. This notion is drawn from the work of Cushman (2008) who presented participants with vignettes that described potentially harmful behavior. Across the conditions, he manipulated the extent to which a character desired to cause harm and the extent to which his actions actually resulted in harm. The character's desire to cause harm (regardless of whether or not it occurred) had the greatest effect on character inferences. The actual harmful consequences of the perpetrator's behavior, however, had the greatest effect on the extent to which the agent's actions warranted punishment.

Putting these findings in the context of the fictive pass asymmetry, it is plausible that fictional code violations are seen as immoral and as a cue to a bad character, to a greater extent than they are seen as deserving of punishment because they do not have any *real* consequences. Based on this, Experiment 4 tested whether participants' desire to punish the vignettes' main character would respond more strongly to the presumption of negative consequences than to inferences of bad character.

Engaging with immoral fiction may not have consequences in the same way that real acts do. Nonetheless, it has been mentioned how a possible consequence of engaging with immoral fiction, specifically that which encourages its consumer to imagine purity code violations, is that it

serves as a signal of one's bad character. It is, however, equally plausible that immoral fiction is seen as being a corruptive influence that can actually inspire bad character and bad behavior.

A cultural mentality that reflects this notion can be seen in the Hayes Code of 1930. The Hayes Code was established by American film production and distribution companies in order to regulate, if not outright prevent, content that they deemed immoral from entering films. The Hayes Code argued that film had the ability to "affect the moral standards of those who, through the screen, take in the ideas and ideals" and "inspire others with a desire for imitation" (Bynum, 2006). Today, similar sentiments are mirrored by tabloids (e.g. Bates & Pow, 2013) and social scientists alike (e.g. Bushman, Gollwitzer, & Cruz, 2015) as they express concerns that media consumers can emulate the behavior that they see on screen (or in the case of a video game, perform), and also become morally corrupted by it.

Of course, moral corruption and behavioral imitation are not mutually exclusive so in the "fiction" conditions of Experiment 4, questions were included that explicitly asked participants the extent to which the fictional acts that were described in the vignettes would worsen the main agent's moral character, and the extent to which he might emulate the behavior. These questions were omitted from the "reality" conditions because they would come across as nonsensical and run the risk of ceiling effects.

Method

Participants

The data of 312 United States residents (195 male; M_{age} = 31.13; SD_{age} = 9.30) were collected from Amazon's Mechanical Turk service. Twelve participants (3.8%) were excluded for failing an attention checking question. As before, anyone who had participated in any of the former experiments or pretests was unable to participate.

Design and Materials

Experiment 4 employed a 2 (code of violation: harm vs purity) x 2 (context of violation: reality vs fiction; collapsed from three levels) that was identical to Experiment 3. As before, the anger and the disgust items consisted of three word-item measures of each emotion (angry, furious, outraged; disgusted, sickened, revolted) and facial expression agreement items (Beaupré, et al., 2000). The character-based judgments were the same as in Experiments 2 and 3. The act-based judgments were largely similar to Experiments 2 and 3, but the phrasing of each item was slightly changed in the fictional contexts. This change was to ensure that participants understood that they were to evaluate the fictional act, not its real-life equivalent. This change was implemented because upon reviewing the act-based measures of the previous experiments, it became apparent that there may have been ambiguity surrounding an item such as, "*Is this morally wrong?"* when it was asked in regards to an act committed in a fictional context. To resolve this, the phrasing of the act-based judgments were adapted to each fictional context (e.g. *Is it morally wrong to* [*imagine*

this/watch this in a film/ do this in a video game]?). It is possible that the consistently high correlation between act and character-based judgments was due to confusing measures. Clarifying the wording of the act-based items should help to reduce this variance.

A second innovation of Experiment 4 is the addition of measures of future consequences. These new items assessed the extent to which participants perceived fictional acts to cause one to become corrupt (a.k.a. they served as a cause of bad character) and the extent to which the acts may be committed in real life. All items were measures on a 1 (*definitely disagree*) to 7 (*definitely agree*) scale, all items were coded so that higher number equaled more consequences. See all of these items below on Table 3.

Despite the categorical differences between these two scales, all nine items collapsed into a single measure of consequences. While these two scales may seem unique, prior research has suggested that one's moral character can signal future behaviors (e.g. Tannenbaum et al., 2011, Experiment 2; Pizarro, et al., 2012) and sure enough, the nine items had a Chronbach's alpha of 0.89. If the items were separated into their respective subscales, a high correlation of r = .89 was revealed, thus supporting the initial decision to analyze the two scales as a single measure. As demonstrated in Table 3, the phrasing of these items was changed slightly between contexts and to reiterate, these questions were only presented to participants who were randomly assigned to evaluate a fictional act.

Table 3

Act and character based measures of consequences, Experiment 4.

Character consequences	Behavioral consequences
Will [imagining these sorts of things/watching these sorts of films/playing these sorts of video games] make Sam a morally bad person?	Does [having thoughts/watching films/playing video games] like this encourage people to do bad things?
Do [thoughts/films/video games] like this corrupt people's character?	Will Sam do this in real life because he [imagined doing it/watching it in a film/ did it in a video game]?
Can these sorts of [thoughts/films/video games] make a good person turn bad?	Do [thoughts/films/video games] like this encourage negative behavior?
This is just a [thought/film/video game] so it won't have any effect on someone's moral character. (Reverse coded)	[Having these thoughts/watching this film/playing this video game] will not cause any real-life consequences. (Reverse coded)
	The fact that Sam is [having these thoughts/watching this film/playing this video game] does not tell us anything about how he will act in real-life. (Reverse coded)

Lastly, participants' *desire to punish* the main character of the vignette was measured with a single item that simply asked the extent to which the individual should be punished for his actions. Like the moral judgments and emotion measures, this item was presented across all contexts both real and fictional. As previously mentioned, the desire to punish is associated with the amount of harm that one's actions cause (Cushman, 2008). As such, it is expected that this item will help to make further distinctions between judgments of moral character and the consequences of engaging with fictional code violations.

Table 4.

Means, standard deviations, and correlation coefficients (Pearson's r), Experiment 4

	1.	2.	3.	4.	5.	6.	7.
1. Anger							
2. Disgust	0.53						
3. Act-based Judgments	0.61	0.54					
4. Character-based Judgments	0.56	0.56	0.78				
5. Consequences	0.51	0.59	0.75	0.73			
6. Punishment†	0.47	0.28	0.83	0.58	0.50		
7. Moral Wrongness†	0.62	0.59	0.95	0.93	0.78	0.75	
Mean	3.04	4.33	3.22	3.16	3.33	2.38	3.19
SD	1.63	1.88	1.88	1.61	1.42	1.88	1.65
Chronbach's Alpha	0.94	0.95	0.88	0.96	0.93		

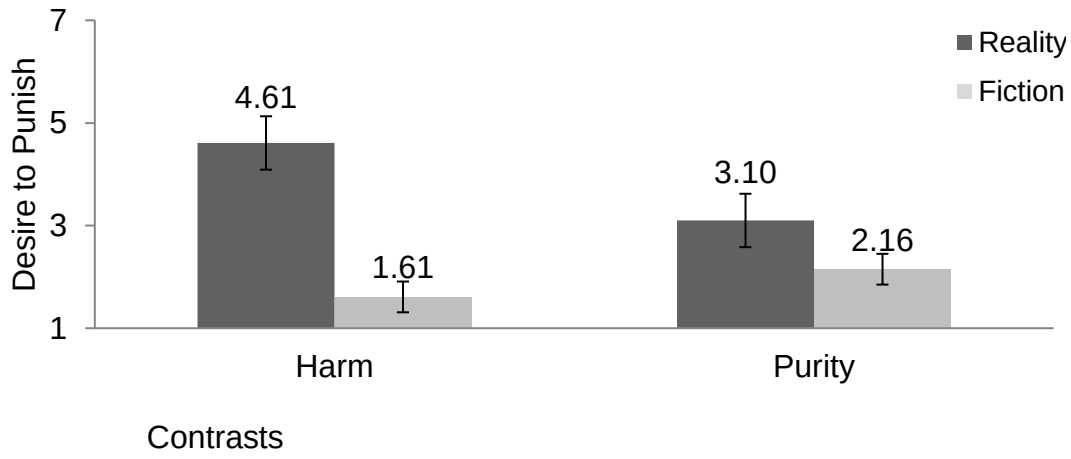
Note. All correlations are significant from zero, $p < .001$; $N = 300$; † = single item scale.

Results

The Fictive Pass Asymmetry

As with the previous experiments, a fictive pass was granted to harm violations more so than it was to purity violations. This was indicated by significant Code x Context interactions for both act and character-based judgments (Figure 7). In other words, harmful acts, compared to impure acts, were relatively more acceptable in fiction than in real-life and unlike in Experiment 3, fictive pass effects were now found for character-based judgments. Simple effect analyses for act-based judgments suggested a relative difference between harm and purity in that both harm code, $F(1, 296)$

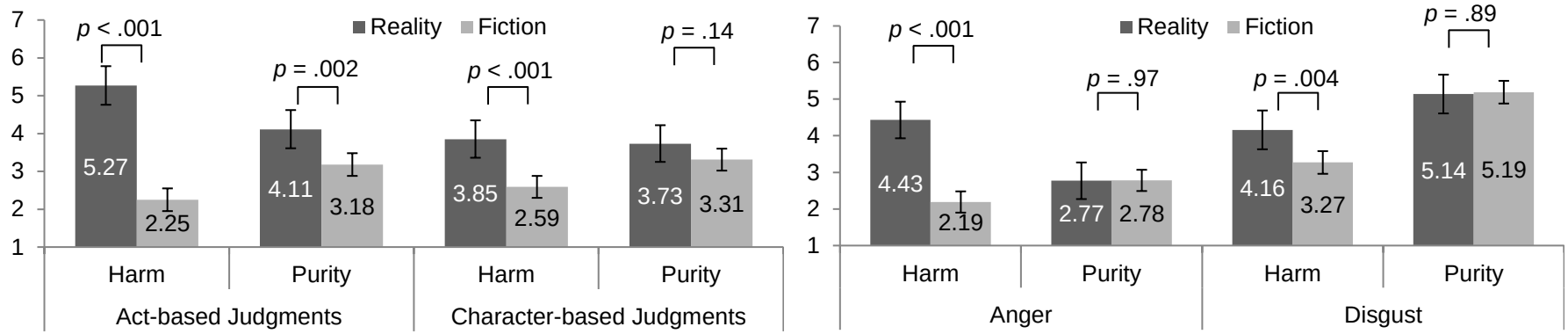
= 100.43, $p < .001$, $\eta_p^2 = .25$. and purity code, $F(1, 296) = 9.70$, $p = .002$, $\eta_p^2 = .03$, violations showed a significant reality to fiction drop, all be it a smaller drop for purity code violations. Simple effects of the character interaction suggested that the reality to fiction difference was more absolute in that that reality to fiction difference was statistically significant for harm violations, $F(1, 296) = 19.09$, $p < .001$, $\eta_p^2 = .06$, but not statistically significant for purity code violations, $F(1, 296) = 2.21$, $p = .14$, $\eta_p^2 = .007$. Punishment ratings also showed a similar fictive pass asymmetry between real and fictional contexts (Figure 6), but this was mostly because real purity code violations, in the absence of any harm, were deemed less punishable than real acts of harm.



Harm vs Purity	Reality vs Fiction	
	Code Main Effect	F (1, 296) = 5.06, p = .03, η_p^2 = .02
	Context Main Effect	F (1, 296) = 85.50, $p < .001$, η_p^2 = .22
	Interaction Effect	F (1, 296) = 23.04, $p < .001$, η_p^2 = .07

Figure 6. Desire to punish across all codes and context, Experiment 4. Means and 95% confidence intervals are presented above. Higher numbers equal more punishment

The moral emotions of anger and disgust showed the expected main effects (Figure 7). Purity violations, unlike harm violations, evoked more disgust relative to anger. Moreover, both emotions displayed a fictive pass asymmetry by showing significant reality to fiction drops for harm code violations, but not for purity code violations. In support of the initial predictions of the fictive pass asymmetry effects on anger and disgust, this asymmetry was less strong for the disgust felt towards the harm condition, as indicated by a simple mixed interaction of Emotion (anger vs disgust) x Context (reality vs fiction), $F(1, 147) = 7.57$, $p = .007$, $\eta_p^2 = .048$. Furthermore, simple effect analyses lent additional support to the fictive pass asymmetry hypothesis by indicating that anger towards harm dropped significantly between contexts, $F(1, 296) = 27.54$, $p < .001$, $\eta_p^2 = 0.09$, but anger towards purity code violations did not significantly change between contexts, $F(1, 296) = .002$, $p = 0.97$. $\eta_p^2 = .00$. The same was true for the disgust, which dropped significantly from reality to fiction for harm violations, $F(1, 296) = 8.20$, $p < .001$, $\eta_p^2 = .03$, but did not change significantly between contexts for purity code violations, $F(1, 296) = .02$, $p = .88$, $\eta_p^2 = .00$.



Contrasts		Dependent Variable	Code Main Effect	Context Main Effect	Interaction Effect
Code	Context				
Harm vs Purity	Reality vs Fiction	Act-based Judgments	$F(1, 296) = .30$, $p = .59$, $\eta_p^2 = .01$	$F(1, 296) = 86.55$, $p < .001$, $\eta_p^2 = .23$	$F(1, 296) = 24.14$, $p < .001$, $\eta_p^2 = .08$
		Character-based Judgments	$F(1, 296) = 2.15$, $p = .14$, $\eta_p^2 = .01$	$F(1, 296) = 17.20$, $p < .001$, $\eta_p^2 = .06$	$F(1, 296) = 4.21$, $p < .05$, $\eta_p^2 = .01$
		Anger	$F(1, 296) = 19.05$, $p < .001$, $\eta_p^2 = .06$	$F(1, 296) = 13.64$, $p < .001$, $\eta_p^2 = .04$	$F(1, 296) = 14.07$, $p < .001$, $\eta_p^2 = .05$
		Disgust	$F(1, 296) = 43.04$, $p < .001$, $\eta_p^2 = .13$	$F(1, 296) = 3.70$, $p < .01$, $\eta_p^2 = .01$	$F(1, 296) = 4.57$, $p = .03$, $\eta_p^2 = .02$

Figure 7. Moral judgments and moral emotions across all codes and contexts, Experiment 4. Means, 95% confidence intervals, and simple effects are presented in chart. Higher numbers equal more condemnation or more negative affect.

Immoral fiction as a cue or a cause of bad character?

A second series of analyses aimed to identify if the greater condemnation of fictional purity, relative to harm, could be explained by the extent to which fictional impurities are viewed as a cue to a pre-existing bad character, or as a cause of future character-related and behavioral consequences. To test this, two analyses of parallel mediation were conducted using the PROCESS macro's fourth model (Hayes, 2012) at 10,000 iterations. For both of these analyses, the predictor variable was a coding of the type of violation (harm vs purity), and the parallel mediators were judgments of moral character (i.e. signaling an already bad character) and the perceived consequences of the act (i.e. a cause of bad character and behavior). For the two analyses, the outcome variables were moral wrongness and the desire to punish. It is important to reiterate that both of these models contained the consequence variable, which was only assigned to the fiction conditions. As such, these analyses excluded the reality condition in favor of specifically examining the effect of character and consequence judgments on the relationship between fictional code violations and moral condemnation.

For both analyses, the two mediating paths were significant (Figure 8), indicating that fictional purity code violations, more so than fictional harm code violations, are condemned because they are seen as a cue to an existing bad character *as well as* a cause of it. However, a comparison of the indirect effects from violation type to moral wrongness showed a statistically significant difference in the strength of these two mediating paths (character minus consequences $b = 0.45$, 95% CI = [0.17, 0.75]), indicating that impure

fiction is seen as a cue to a pre-existing bad character to a greater extent than it is seen as a cause of it. Additional analyses of moderated mediation were conducted that compared the effects of each one-on-one fictional contrast (i.e. imagination vs film; film vs video game). The results, however, indicated that these effects were not further significantly moderated by the specific fictional context in which they occurred.

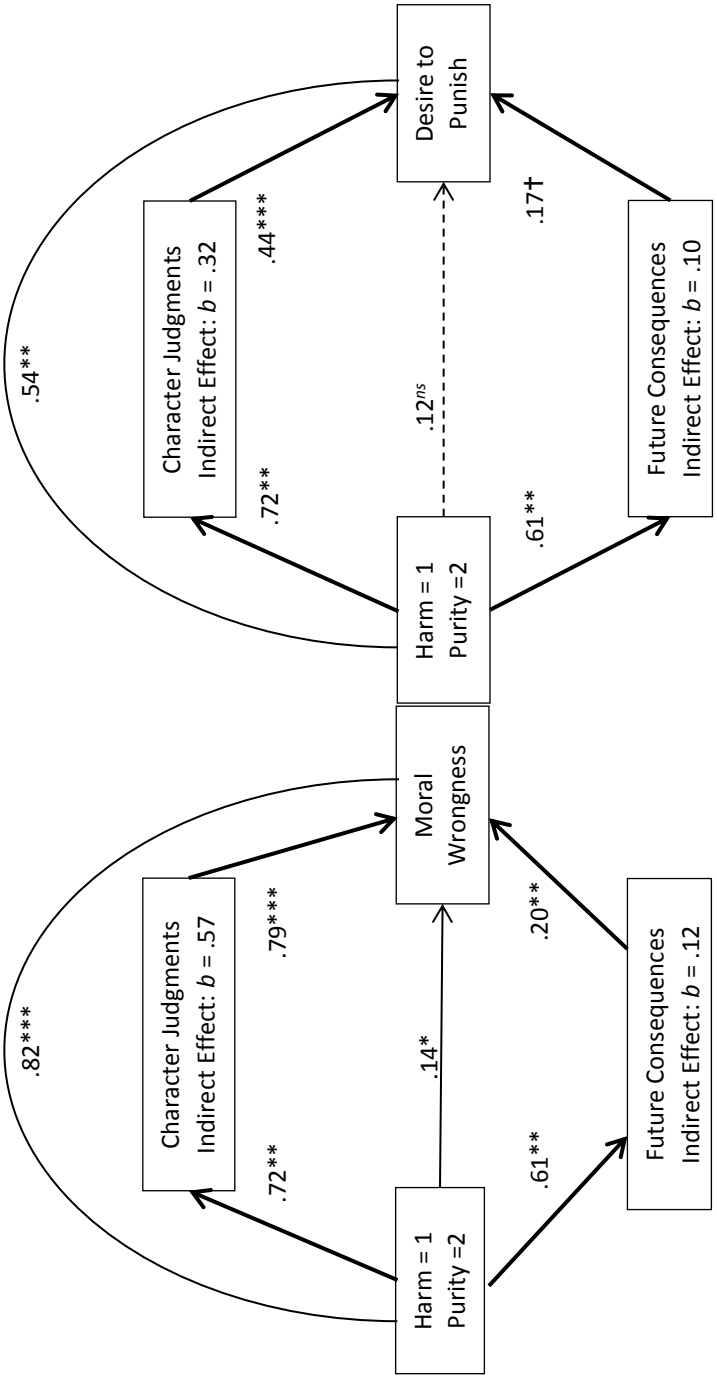


Figure 8. Unstandardized regression coefficients showing mediation of moral character and future consequences of fiction between fictional violations type and moral wrongness, Experiment 4.

Note: † < .10; * < .05; ** < .01; *** < .001. Bold paths indicate significant indirect effects (95% CIs exclude zero). Dashed paths are non-significant.

Discussion

Experiment 4 lent its support to the fictive pass asymmetry hypothesis by providing further evidence that moral condemnation of harm code violations, more so than purity code violations, is mitigated by fictional contexts. Similarly, the amount of emotion that was evoked from violations of harm code violations, but not purity code violations, was mitigated by fictional contexts, and in a similar vein, the effects of the fictive pass asymmetry were less strong for anger than disgust.

A mediation analysis established that impure fiction, more so than harmful fiction, is indicative of a preexisting poor character (i.e. negative character judgments) to a greater extent than it is seen as a cause of bad future character behaviors (i.e. future consequences). Both of these variables were significant predictors of overall moral wrongness, but the effect of character judgments, compared to future consequences, was relatively stronger.

Experiment 5

The results of Experiment 4 offered additional information about the mechanics that drive the fictive pass asymmetry. Results indicated that the consumption of fictional code violations reveals a preexisting bad moral character to a significantly greater extent than it is seen as being the cause of bad character and behaviors. Measures of perceived future consequence were, however, only gathered from participants who were randomly assigned to evaluate fictional code violations. Because the perceived consequences of

real acts were not measured, the mediation analysis of Experiment 4 was only able to offer an explanation for the effect of code violation and it could not divulge information about the full Code x Context interaction on moral wrongness.

Experiment 5 aimed to resolve this by being a direct replication of Experiment 4, except with additional consequence items for the reality condition. These items were previously excluded because by their very nature they are bound to come across as tautological or nonsensical, and are most likely to lead to ceiling effects. In the present experiment, the consequence items were rewritten slightly so that they could be adapted for use in the context of reality while still maintaining clarity. For example, in the fiction conditions two of the items read, "*Does [having thoughts like this/watching this sort of film/playing this sort of video game] encourage people to do bad things?*" and *"Can these sorts of [thoughts/films/video games] make a good person turn bad?"* and in the reality condition these same items were reworded as, "*Do these actions encourage people to do bad things"* and *"Can these sorts of behaviors make a good person turn bad?"*. These questions may come across as slightly redundant. They are, after all, being asked in regards to one who has been explicitly described as committing these acts. Nonetheless, they should be able to assess the extent to which participants presume that the moral code violations will result in future behavioral and character-related consequences. The inclusion of these items will permit a test of mediated moderation that will allow for a more complete understanding of the Code x Context interaction on moral wrongness.

Method

Participants

The data of 352 United States residents (195 male M_{age} = 33.98; SD_{age} = 10.51) collected from Amazon's Mechanical Turk. Four participants (1.1%) were excluded for failing an attention checking question. Participants from any of the former experiments were unable to access the present experiment because of similar measures and manipulations.

Design and Materials

Experiment 5 employed a 2 (code violation: harm vs purity) x 2 (context of violation: reality vs fiction; collapsed from three levels) between-subjects design. The manipulations were entirely unchanged from Experiment 4. The dependent measures were mostly unchanged from Experiment 4. The only change was the previously described addition of future consequence items to the reality condition. The scales that measured anger, disgust, act judgments, character judgments, and consequences were all reliable and they were transformed into composite items. The descriptive statistics of these items can be seen below on Table 5.

Table 5.

Means, standard deviation, correlation coefficients (Pearson's r), and Chronbach's alpha, Experiment 5

		1.	2.	3.	4.	5.	6.	7.
1.	Anger							
2.	Disgust	0.46						
3.	Act-based Judgments	0.59	0.48					

4.	Character-based Judgments	0.58	0.47	0.68				
5.	Consequences	0.52	0.41	0.75	0.70			
6.	Punishment†	0.55	0.31	0.72	0.58	0.60		
7.	Moral Wrongness†	0.57	0.44	0.90	0.68	0.73	0.71	
	Mean	3.03	4.52	3.53	3.28	3.78	2.40	3.40
	SD	1.62	1.81	1.99	1.66	1.45	1.87	2.13
	Chronbach's α =	0.93	0.94	0.88	0.96	0.93		

Note. All correlations are significant from zero, $p < .001$; $N = 352$; † = single item scale.

Results

The Fictive Pass Asymmetry

Experiment 5 successfully replicated the results of Experiment 4. Significant Code x Context interactions for act and character-based moral judgments showed that fictive pass asymmetry effects were stronger for harm code violations than for purity code violations; the reality to fiction drop in condemnation was significantly greater for acts of harm than for impure acts (Figure 9). Simple effects indicated that act-based judgments of harm declined significantly from reality to fiction, $F(1, 343) = 90.31$, $p < .001$, $\eta_p^2 = .21$, as did act-based judgments of impure acts, $F(1, 343) = 5.202$, $p = .02$. $\eta_p^2 = .02$, though the decline for impure acts was relatively less substantial than for harmful acts. Simple effects of the character judgment interaction indicated that harmful acts evoked significantly less sever judgments in fiction than in realty, $F(1,344) = 22.14$, $p < .001$, $\eta_p^2 = .06$, but that character judgments of impure acts did not significantly differ across contexts, $F(1,$

344) = 1.38, p = .25, η_p^2 = .004. Also in line with Experiment 4, overall fictive pass effects were found in participants' desire to punish.

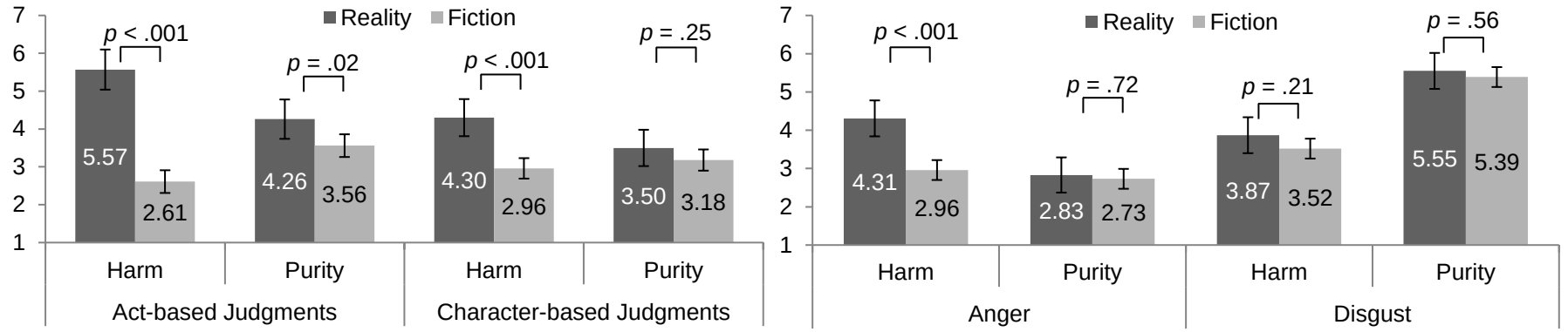
The moral emotions of anger and disgust showed the expected main effects; disgust was more strongly associated with purity code violations and anger was more strongly associated with harm code violations. Moreover, anger, but not disgust, showed the Code x Context interaction that was indicative of fictive pass asymmetry effects. There was no significant difference in the disgust evoked from real versus fictional acts for both harm and purity code violations, thus supporting the hypothesis that the asymmetry would be less pronounced for measures of disgust relative to anger (Figure 9). These findings were qualified by simple effects of the anger interaction, indicating that the anger felt towards harm dropped significantly from reality to fiction, F (1, 344) = 24.31, p < .001, η_p^2 = 0.07, but did not significantly change between contexts for purity violations, F (1, 344) = 0.13, p = .72, η_p^2 = 0.00. Simple effects of the disgust interaction indicated no significant difference between the disgust felt towards real versus fictional acts for either code violation, thus providing further evidence for disgust as an inflexible emotion that is not easily mitigated by contextual information.

The addition of consequence measures across the entire design now allowed them to be tested for fictive pass asymmetry effects. A significant Code x Context interaction indicated that there were fictive pass asymmetry effects for future consequences. Considering that both types of fiction were seen as equally consequential this effect is likely due to the fact that real purity was seen as less consequential than real harm (Figure 10).

Immoral fiction – A cue to bad character or a cause of it?

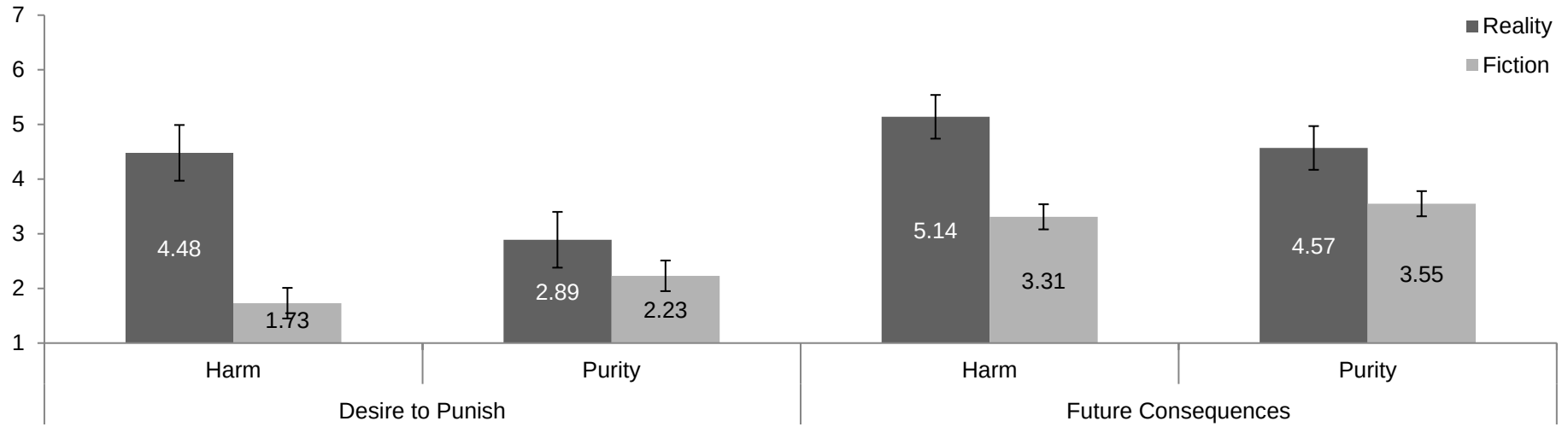
A second series of analyses were conducted to test the mediating effect of moral character and future consequences on the key Code x Context interaction on moral wrongness via mediated moderation. The first analysis was conducted with model 8 of the PROCESS macro (Hayes, 2012) at 10,000 iterations. A coding of the context variable (reality vs fiction) was the model's predictor and a coding of the code variable (harm vs purity) was the moderator. Moral wrongness was the outcome variable and judgments of moral character and future consequences were both placed as parallel mediators. Figure 11 displays a visualization of this model along with its unstandardized regression coefficients.

Indirect effects revealed that the effects of the Code x Context interaction on moral wrongness was significantly mediated by both character judgments (b = 0.38 SE = 0.18, 95% CI = [0.08, 0.79]) and concerns of future consequences (b = 0.51, SE = 0.17, 95% CI = [0.20, 0.89]). Moreover, the conditional direct effects indicated that relationship between purity code violations and moral wrongness was fully explained by the mediators (b 0.05, p = 0.81), but the same was not true for violations of harm (b = -1.31, p < .001).



Contrasts		Dependent Variable	Code Main Effect	Context Main Effect	Interaction Effect
Code	Context				
Harm vs Purity	Reality vs Fiction	Act Judgments	$F(1, 343) = .66$, $p = .42$, $\eta_p^2 = .001$	$F(1, 343) = 70.13$, $p < .001$, $\eta_p^2 = .17$	$F(1, 343) = 26.79$, $p < .001$, $\eta_p^2 = .07$
		Character Judgments	$F(1, 343) = 2.14$, $p = .15$, $\eta_p^2 = .006$	$F(1, 343) = 17.26$, $p < .001$, $\eta_p^2 = .05$	$F(1, 343) = 6.38$, $p = .01$, $\eta_p^2 = .02$
		Anger	$F(1, 343) = 19.69$, $p < .001$, $\eta_p^2 = .05$	$F(1, 343) = 14.06$, $p < .001$, $\eta_p^2 = .04$	$F(1, 343) = 10.56$, $p < .001$, $\eta_p^2 = .03$
		Disgust	$F(1, 343) = 83.68$, $p < .001$, $\eta_p^2 = .20$	$F(1, 343) = 1.68$, $p < .20$, $\eta_p^2 = .005$	$F(1, 343) = .26$, $p < .61$, $\eta_p^2 = .001$

Figure 9. Moral condemnation and moral emotions across all codes and contexts, Experiment 5. Means, 95% confidence intervals, and simple effects are presented in chart. Higher numbers equal moral condemnation or negative emotion.



Contrasts		Dependent Variable	Code Main Effect	Context Main Effect	Interaction Effect
Code	Context				
Harm vs Purity	Reality vs Fiction	Desire to Punish	$F(1, 343) = 6.90$, $p < .01$, $\eta_p^2 = .02$	$F(1, 343) = 66.89$, $p < .001$, $\eta_p^2 = .16$	$F(1, 343) = 25.24$, $p < .001$, $\eta_p^2 = .07$
		Future Consequences	$F(1, 343) = 1.01$, $p = .32$, $\eta_p^2 = .003$	$F(1, 343) = 77.36$, $p < .001$, $\eta_p^2 = .18$	$F(1, 343) = 6.26$, $p < .05$, $\eta_p^2 = .02$

Figure 10. Desire to punish and future consequences across all codes and contexts, Experiment 5. Means and 95% confidence intercals are show above. Higher numbers equal more punishment or greater future consequences.

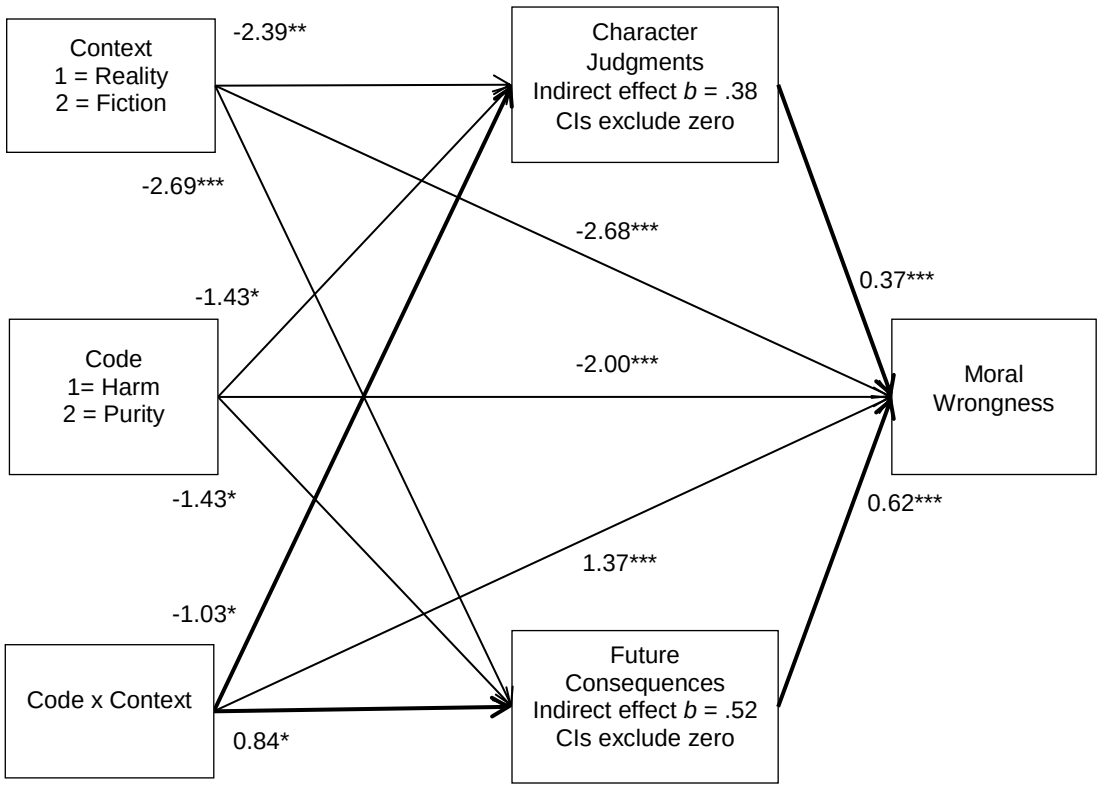


Figure 11. Unstandardized regression coefficients showing the mediating effects of character judgments and future consequences on the interactive effect of Code x Context on moral wrongness, Experiment 5. Note: * < .05; ** < .01; *** < .001. Both paths indicate significant indirect effects (95% CIs exclude zero).

Following the lead of Experiment 4, a second analysis of mediated moderation was conducted with punishment ratings as the outcome variable. Similar to the former analysis on moral wrongness, the inclusion of the reality condition into the model allowed the analysis to now reveal that the effects of the Code x Context interaction on punishment were significantly mediated by both character judgments (b = 0.38, SE = 0.18, CI = [0.08, 0.79]) and

concerns of future consequences (b = 0.51, SE = 0.17, CI = [0.21, 0.89]). Furthermore, the conditional direct effects showed that the relationship between purity code violations and punishment was entirely explained by the character and future consequence mediators (b = -1.31, SE = 0.23, p < .001), but that the relationship between harm violations and punishment was not entirely accounted for (b = 0.05, SE = 0.22, p = .81). It should, however, be noted that moral wrongness and punishment had a strong correlation of r = .72, p < .001, so it is not surprising that they showed similar effects. Experiment 5's inclusion of the reality condition into the model of mediated moderation has revealed that the variance in the effects of the fictive pass asymmetry can be explained by the fact that in fictional contexts, purity code violations, as much as harm code violations, are seen as both a cue to a pre-existing bad character and as a cause of future behavioral and character-related negative consequences.

Discussion

Experiment 5 successfully replicated, and expanded on, the findings of the former experiments. In support of the general fictive pass asymmetry hypothesis, harm, relative to purity code violations, showed a significantly larger gap between evaluations of real versus fictional acts. Innovations of Experiment 5, however, allowed for analyses of moderated mediation to provide a deeper understanding of this effect.

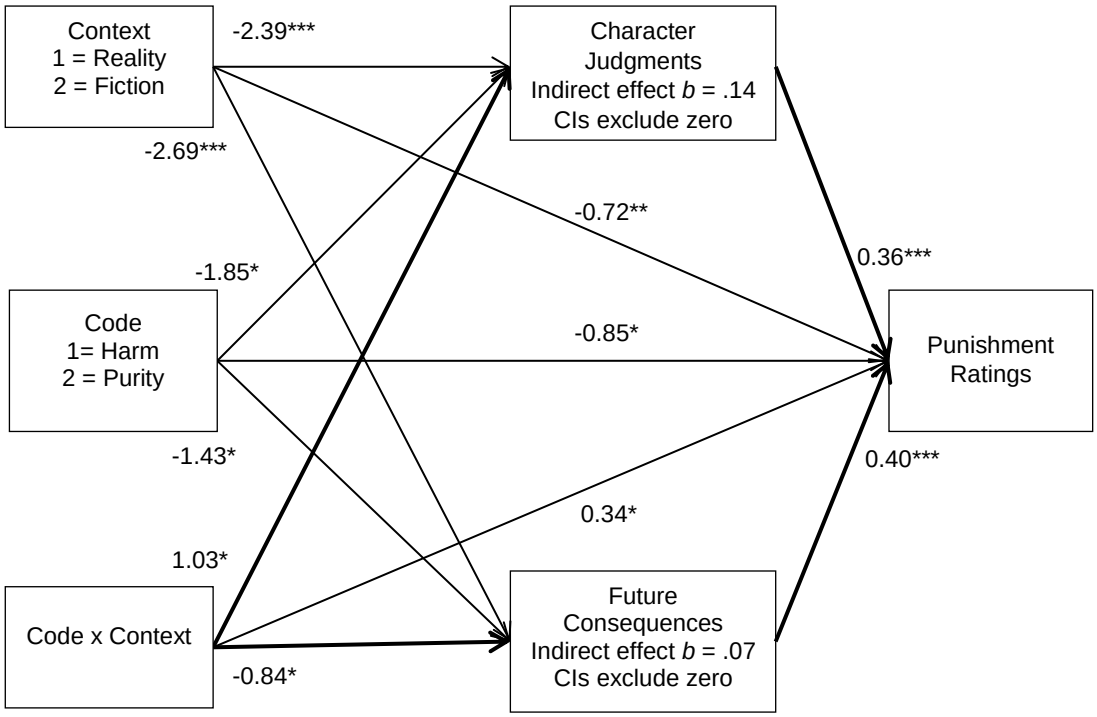


Figure 12. Unstandardized regression coefficients showing the mediating effects of character judgments and future consequences on the interactive effect of Code x Context on punishment ratings, Experiment 5. Note: * < .05; ** < .01; *** < .001. Both paths indicate significant indirect effects (95% CIs exclude zero).

In the previous experiment, a series of simple mediation analyses suggested that violations of fictional purity, relative to harm, indicate bad character to a significantly greater extent than they are perceived as causing it because of downstream behavioral and character-related consequences. These findings offered insight into the evaluative discrepancy between fictional taboos, but they fell short of explaining why the moral condemnation that is directed towards fictional impurities, compared to harms, is relatively closer to the condemnation of their real-life counterparts.

Experiment 5 included measures of future consequences across the whole design and this innovation allowed the above question to be addressed. A more comprehensive analysis of mediated moderation fully explained the effect of the Code x Context interaction on moral wrongness and punishment. Experiment 5 replicated and expanded upon the findings of Experiment 4 by revealing that purity violations, to a greater extent than harm violations, are seen as a cue to bad character more than a cause of it. Moreover, analyses of mediated moderation confirmed that this was true regardless of the real or fictional context in which the act occurred.

Conclusion

The two experiments that are reported in this chapter have successfully replicated the basic Code x Context interaction effects of the fictive pass asymmetry hypothesis. These experiments have also revealed novel information about the role that moral character plays in forming the observed discrepancy in moral wrongness between real and fictional violations of harm and purity. Experiment 4 provided evidence that the consumption of impure fiction, more so than harmful fiction, is indicative of poor moral character to a significantly greater extent in that it raises concerns about future behavioral or character-related consequences. Experiment 5 expanded upon these findings by showing that purity code violations, compared to harm code violations, signal bad character regardless of the context in which they occur.

CHAPTER 6

ALTERNATIVE EXPLANATIONS AND A META-ANALYSIS OF THE FICTIVE PASS ASYMMETRY EFFECTS

The previous empirical chapters have consistently identified novel differences in the extent to which moral violations of harm and purity are condemned across real and fictional contexts. The experiments in chapter 4 explained the development and the refinement of the methodology that is now able to reliably identify fictive pass asymmetry effects. Then chapter 5 dug deeper into the fictive pass asymmetry and demonstrated how the reality to fiction gap in moral condemnation was relatively smaller for purity code violations than for harm code violations because the relationship between violations of purity, both real and fictional, and moral condemnation is uniquely explained by negative character judgments. The present chapter consists of two parts: a final experiment that will aim to replicate and further expand on the fictive pass asymmetry and a meta-analysis of a select number of the experiments in this thesis.

The final experiment aims to address potential theoretical concerns and to refine a couple of measurements. Firstly, the present experiment will explore the effects of the fictive pass asymmetry in relation to an alternate perspective on the definition of the moral codes (Cameron, Lindquist, & Gray, 2015). Secondly, the important role of character in the fictive pass asymmetry effects was demonstrated in chapter 5, so the present experiment will more thoroughly acknowledge the multifaceted nature of the construct of moral character (Goodwin et al., 2014). Thirdly, the emotion measures were slightly refined in an attempt to make anger and disgust as

distinct as possible. Lastly, the chapter will conclude with a series of meta-analyses. It is acknowledged that across these experiments the results have been somewhat inconsistent, and some experiments found effects where other did not. The meta-analyses will attempt to demonstrate how these inconsistencies are mere hiccups and that the fictive pass asymmetry effects will present themselves when the results of these experiments are examined holistically.

Experiment 6

As previously described, a number of innovations will be implemented in Experiment 6 in order to increase the validity and generalizability of the fictive pass asymmetry effects. The below sections will describe these changes in greater detail.

New Vignette Selection Criteria

For Experiment 6, a new set of vignettes will be pretested. These new manipulations will aim to generalize the effects of the fictive pass asymmetry across a greater variety of morally relevant situations. More importantly, however, they should address potential theoretical and methodological concerns of the vignettes that were used in the former experiments.

The vignettes that were used in Experiments 3 – 5 were selected based on their ability to evoke more anger than disgust for violations of harm, and vice versa for violations of purity. This selection method is empirically supported by a sizeable body of evidence that has identified these specific links between moral content and emotional reactions (i.e. Goldberg, et al., 1999, Horberg et al., 2009, Piazza, et al., 2013, Russell & Piazza, 2014).

These links have, however, been refuted and argued against (Cameron, et al., 2015), thus questioning the validity of the former method of vignette selection.

Secondly, despite the best efforts of pretesting, the former experiments showed a difference in moral wrongness between harmful and impure acts in the reality condition. The moral wrongness of real purity was lower than that of harm. This inequality was so prominent in Experiment 2 that floor effects of the purity condition presented themselves as a possible explanation of the fictive pass asymmetry effects. This possibility was, however, dismissed through additional analyses. In Experiments 3 – 5 the moral inequality between real code violations was still present. In spite of this, the moral wrongness of real acts was near the scale midpoint which means that floor effects were not responsible for the findings of these experiments'. Nonetheless, it would be most ideal if real code violations were morally equivalent to one another because this would allow for the clearest expression of fictive pass effects.

The pretest of Experiment 6 will attempt to resolve these two issues by obtaining vignettes via different criteria. Rather than identifying harm and purity code violations based on the extent to which they evoke varying amounts of anger and disgust, the pretest of Experiment 6 will identify these acts based on the extent to which they are seen as immoral and harmful to other people, versus immoral but not harmful to other people. A more detailed discussion of the theory that surrounds this decision will be discussed in the below section *The Perceived Harm in Harmless Impurities*. In short, however, this simple test should successfully select vignettes that

are able to demonstrate that fictive pass asymmetry effects occur even when the moral violations of harm and purity are selected based on the criteria of a theoretical framework which claims that the condemnation of objectively harmless purity violations is rooted in harm concerns (Cameron et al., 2015). This new vignette selection process will validate the former selection method and satisfy the alternate theoretical perspective that acts without harm are condemned because they are seen as harming entities beyond other people such as the self or nature (Cameron et al., 2015; Gray, Waytz, & Young, 2012; Guttierez & Giner-Sorolla, 2011).

Metaphors as Emotion Measures

A second innovation of Experiment 6 will be to make additional efforts to address the traditionally large amount of variance that is shared by anger and disgust (Russell & Giner-Sorolla, 2011a). The variance shared between these emotions has been steadily reducing across the experiments. Nonetheless, it is possible that the metaphorical use of disgust language, as an expression of moral anger (Russell, et al, 2013), has not been entirely accounted for by using the facial expression agreement measures. To address this, the emotion measures of Experiment 6's pretest, and Experiment 6 itself, will include metaphors of physical disgust (i.e. *This makes me want to gag*) that focus on the oral inhibitory aspects of the emotion (Royzman, Atanasov, Landy, Parks, & Gepty, 2014). For parallelism, metaphors of anger will also be added (i.e. *This makes my blood boil*). Expanding upon the emotion measures should allow for further empirical distinction between these two emotions.

The Perceived Harm in Harmless Impurities

As discussed in the second chapter, a large body of research has experimentally demonstrated how morally relevant acts can be partitioned into harm and purity code violations, amongst others, to form distinct moral domains (Chakroff, et al., 2013; Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011; Rozin, et al., 1999; Russell, et al., 2013). The experiments reported in this thesis, with the exception of Experiment 1, have made this distinction as clear as possible by keeping violations of harm free of physical disgust, and by keeping violations of purity free of harm. Despite these efforts, an alternate view may challenge the extent to which these two moral codes are theoretically distinct from one another. Most notably, it has been argued that laypersons perceive harm in objectively harmless purity code violations (Gray, Waytz, & Young, 2012), and that researchers' intuitions about harm can dramatically vary from that of participants.

To contextualize this alternate perspective, consider Haidt's (2001) classic example of an objectively harmless purity code violation that describes two siblings who decide to have sex with each other. This act is explicitly described as a consensual, safe (i.e. extensive use of birth control), and enjoyable one-off experience. The siblings never told anyone that they had done this and this secret made them feel closer and it improved their sibling bond.

Many would agree that this exemplifies an impure, yet objectively harmless act (Chakroff & Young, 2015; Haidt 2001; Guttierez & Giner-Sorolla, 2007; Giner-Sorolla et al., 2012), but an alternative perspective

proposed by Gray and Wegner (2010) would categorize this act as still causing harm, just to non-typical targets. Like many moral psychologists, Gray and Wegner (2010) agree that a prototypically harmful act involves some sort of injury, physical or otherwise, to another specific individual. Unlike others, however, they are not entirely satisfied with the categorization of an objectively harmless "purity" code violation, such as that above example. They argue that such acts *do* involve harm because people see them as causing harm to an abstract or non-specific entity such as God or the natural order (Gray, Schein, & Ward, 2014; Gray, Waytz, & Young, 2012). Gray et al., (2014) argue this stance by explaining how people have a dyadic template of morality that consists of an offender and a victim. In the absence of an explicit victim, people will still rationalize the occurrence of harm by directing it towards the aforementioned non-specific entities. In effect, this blurs the lines that the moral foundations theory (Graham et al., 2011) draws between violations of harm and purity and questions the validity of the vignettes that have been used in the former experiments.

In light of these controversies within moral psychology, Experiment 6 will be sensitive to this alternative perspective on moral violation categorization. Experiment 6 will test the possibility that the acts that have been previously labeled as "harm" code violations can also be characterized as specifically "interpersonal harm" while the formerly labeled "purity" code violations can be characterized as acts that harm other entities such as the self (Chakroff, et al., 2013) or the natural order of things (Gutierrez & Giner-Sorolla, 2011). To initially test this, the pretest will contain a simple item that measures how much harm the acts cause to other people. In Experiment 6

itself, new items will be introduced that will measure the perceived *harmfulness* of each code violation to a variety of targets. It is predicted that this will reinforce the prior classification of harm versus purity by showing a clear difference between them based on the nature of the presumed harm.

A More Heterogeneous Assessment of Moral Character

All of the previously reported experiments have measured moral character with relatively simple items such as, "*Is [Agent] rotten inside?"* and "*Is [Agent] mainly a good person or mainly a bad person?*" Items, such as the former of these two examples, may have confounded moral character with disgust concepts while the latter may have overlooked the heterogeneous nature of moral character. Evaluations of moral character go beyond simple evaluations of good versus bad because specific prosocial traits are also relevant to one's character (Goodwin, et al., 2014). To address this, Experiment 6 included a number of specific traits as character measures. These traits included warmth, fairness, empathy, integrity, and abnormality (Goodwin et al., 2014; Uhlmann, et al., 2015). This innovation will more thoroughly encapsulate the scope of moral character and allow it to be more independent from the disgust construct.

Desire as an Alternate Explanation of the Fictive Pass Asymmetry

The final innovation of Experiment 6 sought to identify explanations of the fictive pass asymmetry beyond the character-based explanations that were established in the experiments of the previous chapter. To achieve this, inspiration was drawn from Russell and Piazza (2015, Experiment 4) who found that bizarre sexual acts were condemned as much as they were

desired as when they were actually acted upon. Considering these findings in relation to the fictive pass asymmetry effects allows for the possibility that the asymmetry may be driven by desire. Perhaps those who consume fictional impurities will be seen as actually desiring to perform the act more so than people who consume fictional harm. This hypothesis will be tested by measuring the extent to which the vignettes' main character is perceived as desiring to commit the described act. The measures of desire will be tested as a mediator of the fictive pass asymmetry effects, as was done with the character items in the previous experiments.

Pretest

A pretest was conducted in order to identify vignettes that could be used in Experiment 6 and to pilot new dependent measures such as the emotional metaphors and a measure of harmfulness. As previously stated, the aims of this pretest were twofold. Firstly, it aimed to identify new vignettes that could generalize the effects of the fictive pass asymmetry. Secondly, it aimed to re-categorize "harm" and "purity" code violations, respectively, as the extent to which they are immoral and cause harm to other people, versus immoral and do not cause harm to other people.

Forty vignettes were written that described morally relevant behavior of varying severity. The "harm code violation" vignettes described acts that caused specific instances of interpersonal harm, contained no elements of physical disgust, and did not involve abnormal uses of the body. The "purity code violation" vignettes described acts that violated bodily-moral norms but were entirely consensual and free of interpersonal harm. All pretested

vignettes were described as occurring in real-life. Emotion measures of anger and disgust consisted of word item measures, metaphorical expressions (i.e. *this makes my blood boil; this makes me want to gag*), and facial expression agreement measures of each emotion. Single items measured moral wrongness (*How morally wrong is this?*) and the presumption of harm to specific others (*How much harm does this cause to other people?*). All dependent measures were on seven point scales from 1 (n*ot at all*) to 7 (e*ntirely*). Data was collected from 128 United States residents though Mechanical Turk. Each participant was randomly presented with eight of the forty vignettes and no demographic information was collected.

From the forty vignettes, four were selected (two of each code violation) to be used in Experiment 6 based on the extent to which they were immoral and harmful to other people (harm code violations) versus equally immoral but not harmful to other people (purity code violations). Collapsing the four selected vignettes into their respective code violations showed that there was no significant difference in how morally wrong the two types of behaviors were, $F(1, 3) = 3.47$, $p = .20$ and that “harm” violations were significantly more harmful to other people than “purity” violations, $F(1, 3) = 125.34$, $p = .008$. Although it was not factored into the selection process, purity code violations evoked significantly more disgust than anger, $F(1,3) = 198.96$, $p = .005$ and both code violations evoked about the same amount of anger, $F(1, 3) = 6.43$, $p = .13$. The selected violations read as follows:

Harm Vignettes

"Sam comments loudly about how fat a woman looks in her jeans. Sam enjoys doing this."

"Sam gets into a fight with another man and he punches the other guy in the face. He enjoys doing this."

Purity Vignettes

"Sam is a member of a cannibal society. Members of this society donate their bodies so that upon death the living members can cook and eat them. Sam enjoys being part of this club."

"Sam has sex with a frozen chicken before cooking it and eating it for dinner. Sam enjoys doing this."

Method

Participants

The data of 484 United States residents (248 male*; M_{age}* = 34.77; *SD_{age} = 10.36*) were collected from Amazon's Mechanical Turk service. The questionnaire contained an attention checking question, hidden amongst the dependent measures, that asked participants to select a specific scale point. If this attention check was failed then the questionnaire was programmed to direct them to the debrief, flag their data as incomplete, and recruit a new participant in their place. As such, all participants that fully completed the questionnaire were included in the final data set.

Design, Procedure, and Materials

Experiment 6 was a 2 (moral code violation: harm vs purity) x 2 (context of violation: reality vs fiction, collapsed from three levels) between-subjects design. The three fictional contexts (imagined, watched in a film, performed in a video game) were collapsed into a single level of fiction. The wording of the vignettes was manipulated to present the acts as occurring in the different contexts. For instance, in the context of imagination, one of the purity vignettes from above was written as, "*Sam imagined 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 Likert-type scales that ranged from 1 (*not at all*) to 7 (*entirely*). Participants were randomly presented with a harm or purity code violation. Further randomization presented it as occurring in one of the four contexts (real life, imagined, watched in a film, performed in a video game). Following the manipulation, participants were presented with the dependent variables, which were randomized, and lastly, participants reported their age and sex before being debriefed.

Emotion Measures

The word-item and the facial expression agreement measures were unchanged from the previous experiments. As previously discussed, an innovation of Experiment 6 was the inclusion of metaphors as measures of each emotion. Anger metaphors included, *this makes my blood boil, this makes me feel like I will lose my cool,* and *this makes me see red*. Disgust metaphors included, *this makes me feel like I will lose my appetite; this makes me want to gag*; and *this makes me sick to my stomach.*

All of the anger and disgust items were collapsed into their respective composite variables. The anger items (α = .95) and the disgust items (α = .96) were both reliable and had a moderate, but statistically significant, correlation (r = 0.50, p < .001) that was low enough to analyze them separately.

Act Judgments

The act-based judgments were unchanged from the experiments in the previous chapters. As in Experiments 4 and 5, the wording of these items was slightly modified for each context. This was to ensure that participants in a fiction condition understood that they were to be evaluating the consumption of the fiction, not its real-life equivalent (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

Based on past research that has explored the heterogeneous nature of moral character (e.g. Goodwin et al., 2014; Uhlmann, Pizarro, & Diermeier, 2015), participants reported the extent to which they perceived the main agent as *abnormal, twisted, 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 and fairness). The items that loaded on the second factor related to negative and abnormal traits (i.e. deviant and twisted). The items that

composed each of these factors had strong reliabilities (both α's = .94) and shared a correlation of $r = .29$, $p < .001$, thus allowing them to be collapsed into two composite variables: moral character and abnormal character. All items were coded so that higher numbers reflected more immorality or abnormality.

Measures of Harmfulness

The perception 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 (e.g. *By [imagining this/watching this film/playing this video game] did Sam cause [psychological/physical /emotional] harm to anyone other than himself?)*.

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?*; α = .90.)

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

Natural Harm: Two items measured the perceived harm the agent's actions caused to the natural order (*Do you think that Sam's behavior*

caused damage to the natural order of things? Did Sam's actions violate any laws of nature?; r = .73).

The items of these three scales were subjected to an exploratory factor analysis with a maximum likelihood extraction and a promax rotation. Results indicated that these items loaded onto two distinct factors. The natural harm and self harm items (5 items: α = .88) loaded on to one factor and were averaged into the variable *non-social harm*. The *social harm* items formed another factor (5 items; α = .90). Even though these items are all rooted in harm concerns, these two factor loadings lend initial empirical support to the idea that acts seen to harm other people, either individually or collectively, form a different class of moral violation than acts that are seen as not harming specific others.

Measures of the Presumption of Desire

Three items (α = .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?*). As shown in the example, the wording of these items was modified to fit the different levels of the context variable.

Results

The Fictive Pass Asymmetry

Main effects of the moral judgments indicated that there was no significant difference in act condemnation between harm and purity code violations. Harm, relative to purity, was more associated with poor character

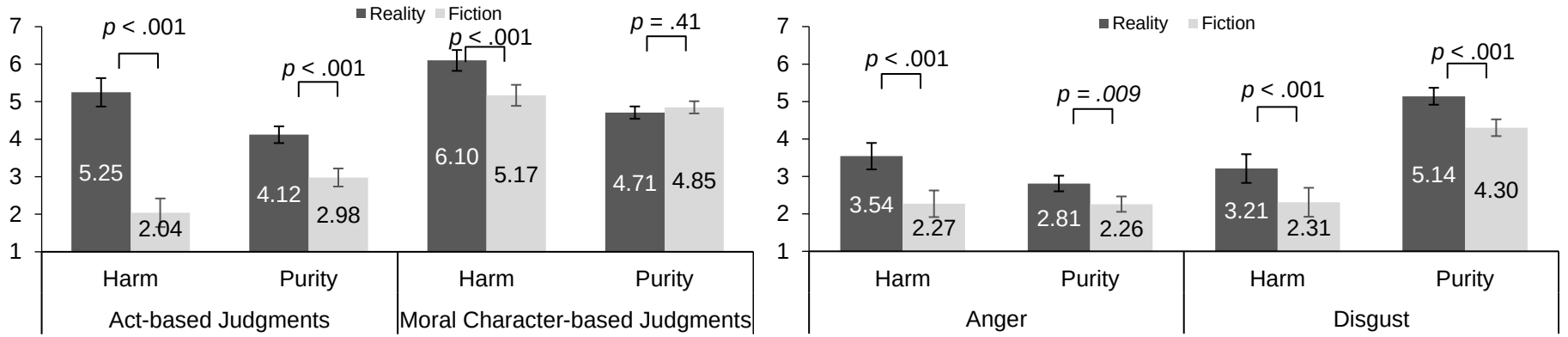
and for both act and character based judgments, real acts were more condemnable than their fictional counterparts.

As in previous experiments, significant Code x Context interactions revealed fictive pass asymmetry effects for act and character-based moral judgments. In other words, the drop in moral condemnation between real and fictional contexts was significantly greater for acts that violated harm norms than it was for acts that violated purity norms (Figure 13). As in previous experiments, the simple effects of the act-based judgment interaction revealed that the fictive pass effects occurred in relative, but not absolute terms. That is, the reality to fiction drop was statistically significant for both harm, $F(1, 450) = 206.83$, $p < .001$, $\eta_p^2 = .32$ and purity, $F(1, 450) = 24.63$, $p < .001$, $\eta_p^2 = .05$ code violations, but the drop was still smaller for the purity code violations, as indicated by the interaction effect as well as the partial eta squared of the simple effects. Simple effects for the interaction of moral character demonstrated the effects of the fictive pass asymmetry in more absolute terms. For judgments of moral character the reality to fiction drop in condemnation was statistically significant for harm code violations, $F(1, 480) = 32.09$, $p < .001$, $\eta_p^2 = .06$, but not statistically significant for purity code violations, $F(1, 480) = 0.69$, $p < .41$, $\eta_p^2 = .001$.

In pretesting, one of the criteria by which vignettes were selected was equivalence in moral wrongness between harm and purity code violations. Notwithstanding, an analysis that included only real acts showed that real harm was significantly more immoral than real impurities $F(1, 122) = 17.75$, $p < .001$, $\eta_p^2 = 0.13$. In spite of this unexpected divergence from the results of

the pretest, an examination of the act-based moral wrongness means (Figure 13) indicates that harm violations may have started higher on the scale than purity violations did, but they also ended at a lower point. Compare this to the purity violations that started at a lower point, but in fictional contexts their wrongness remained relatively closer to the scale midpoint. Effectively, these observations alleviate doubts that the baseline difference in moral wrongness is responsible for the effects of the fictive pass asymmetry because of a floor effect of the purity condition. If anything, the result most indicative of a floor effect is for fictional harm code violations which works in favor of the fictive pass asymmetry by highlighting the relative banality with which people assess fictional harm.

Main effects of anger and disgust showed that harm code violations but not purity code violations showed more anger than disgust and vice versa (Figure 13). Consistent with Experiments 4 and 5 only anger showed fictive pass asymmetry effects but the reality-to-fiction drop for disgust was non-significant for both harm and purity code violations (Figure 13). Simple effects of the interactions of the emotion variables indicated that anger towards harm violations declined significantly from reality to fiction, $F(1, 480) = 37.26$, $p < .001$, $\eta_p^2 = .07$, and that disgust towards harm also declined significantly from reality to fiction, $F(1, 480) = 6.89$, $p = .009$, $\eta_p^2 = .014$. Similarly for the disgust interaction, disgust towards harm showed a significant reality to fiction decline, $F(1, 480) = 16.03$, $p < .001$, $\eta_p^2 = .032$, as did disgust towards impurities, $F(1, 480) = 32.17$, $p < .001$, $\eta_p^2 = .028$.



Contrasts		Dependent Variable	Code Main Effect	Context Main Effect	Interaction Effect
Code	Context				
Harm vs Purity	Reality vs Fiction	Act-based Judgments	$F(1, 450) = 0.36$, $p = .55$, $\eta_p^2 = .001$	$F(1, 450) = 184.52$, $p < .001$, $\eta_p^2 = .29$	$F(1, 450) = 41.82$, $p < .001$, $\eta_p^2 = .09$
		Character-based Judgments	$F(1, 480) = 53.74$, $p < .001$, $\eta_p^2 = .10$	$F(1, 480) = 11.61$, $p = .001$, $\eta_p^2 = .02$	$F(1, 480) = 21.10$, $p < .001$, $\eta_p^2 = .04$
		Anger	$F(1, 480) = 6.39$, $p = .01$, $\eta_p^2 = .01$	$F(1, 480) = 38.03$, $p < .001$, $\eta_p^2 = .07$	$F(1, 480) = 6.00$, $p = .02$, $\eta_p^2 = .01$
		Disgust	$F(1, 480) = 149.93$, $p < .001$, $\eta_p^2 = .24$	$F(1, 480) = 29.64$, $p < .001$, $\eta_p^2 = .06$	$F(1, 480) = 0.04$, $p = .84$, $\eta_p^2 = .00$

Figure 13. Moral judgments and moral emotions across all codes and contexts, Experiment 6. Means, 95% confidence intervals, and simple effects are presented in chart. Higher numbers equal more condemnation or more negative emotion.

Main effects of abnormal character (Figure 14) indicated that real acts, relative to fictional acts, F (1, 480) = 83.74, $p < .001$, $\eta_p^2 = 0.15$, and impure acts, relative to harmful acts, F (1, 480) = 102.81, $p < .001$, $\eta_p^2 = 0.18$, were most strongly associated with character abnormality. The Code x Context interaction was not statistically significant, but as close to marginal significance as the interaction was ($p = 0.11$), character abnormality is a variable that should not be excluded from future analyses and future research.

Main effects of *desire* indicted that purity code violations were more indicative of desires than harm violations, $F(1, 453) = 4.82$, $p = .03$, $\eta_p^2 = 0.25$. Measures of desire did not, however, yield a significant Code x Context fictive pass asymmetry interaction ($p = 0.57$).

The harm of "(impersonally) harmless" purity code violations

To reiterate, the newly added measures of *harmfulness* addressed the possibility that the harm-purity code distinction that has been relied upon in previous experiments can alternatively be described as a distinction between acts that explicitly harm other people and acts that harm a non-social entity (Gray et al., 2014).

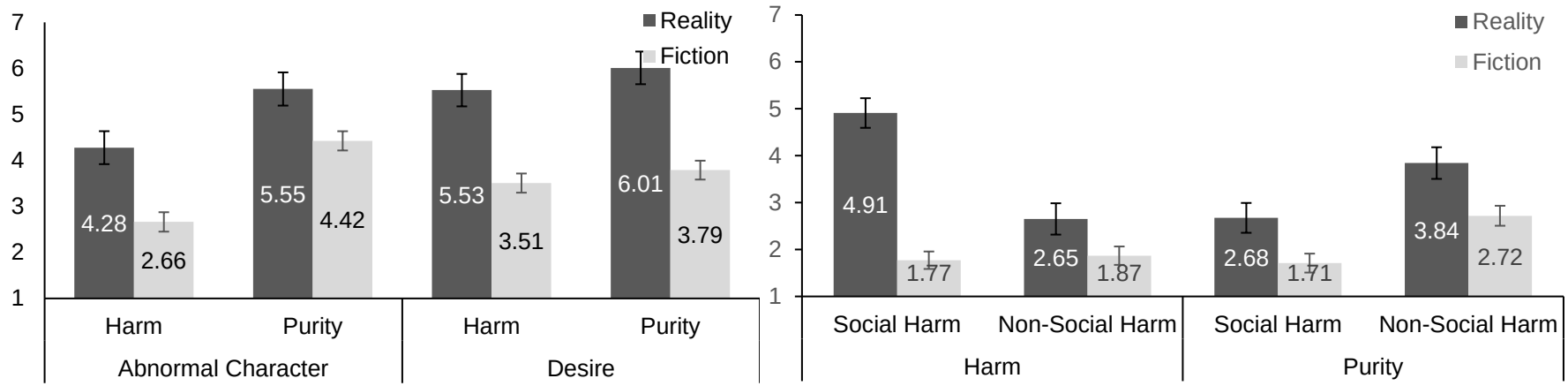
The effect of the manipulations on the perceived type of harm (social versus non-social) was shown in a significant Harm Type X Code Violations X Context interaction, $F(1, 450) = 107.75$, $p < .001$, $\eta_p^2 = 0.19$. More specifically, a Harm Type x Code Violation interaction, $F(1, 450) = 321.01$, $p < .001$, $\eta_p^2 = 0.42$, indicated that harmful acts evoked stronger perceptions of social harms ($M = 3.34$) than non-social harms ($M = 2.26$, $p < .001$), and

impure acts more strongly evoked perceptions of non-social harms (M = 3.28) than social harms (M = 2.19, $p < .001$). This suggests that the previously used harm versus purity distinction can also be characterized in terms of acts that harm social versus non-social entities, thus displaying unique differences between these moral codes and addressing theoretical concerns about their homogeneity (Cameron et al., 2015).

In further support of this distinction, the different types of harm were also associated with different moral emotions. When controlling for disgust, social harm (b = 0.37, $p < .001$) more so than non-social harm (b = 0.05, p = .32) predicted anger; when controlling for anger, non-social harm was a strong positive predictor of disgust (b = 0.71, $p < .001$), 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 15 for a visualization of this model as well as the unstandardized regression coefficients.



Contrasts		Dependent Variable	Code Main Effect	Context Main Effect	Interaction Effect
Code	Context				
Harm vs Purity	Reality vs Fiction	Abnormal Character	$F(1, 480) = 120.81$, $p < .001$, $\eta_p^2 = .18$	$F(1, 480) = 83.74$, $p < .001$, $\eta_p^2 = .15$	$F(1, 480) = 2.64$, $p = .11$, $\eta_p^2 = .005$
		Desire	$F(1, 450) = 4.82$, $p = .03$, $\eta_p^2 = .10$	$F(1, 450) = 147.29$, $p < .001$, $\eta_p^2 = .25$	$F(1, 450) = 0.31$, $p = .57$, $\eta_p^2 = .001$
		Social Harm	$F(1, 450) = 73.52$, $p < .001$, $\eta_p^2 = .14$	$F(1, 450) = 234.67$, $p < .001$, $\eta_p^2 = .34$	$F(1, 450) = 65.81$, $p < .001$, $\eta_p^2 = .13$
		Non-Social Harm	$F(1, 450) = 52.15$, $p < .001$, $\eta_p^2 = .10$	$F(1, 450) = 45.42$, $p < .001$, $\eta_p^2 = .09$	$F(1, 450) = 1.45$, $p = .23$, $\eta_p^2 = .00$

Figure 14. Abnormal character judgments, desire, harmfulness across all codes and contexts, Experiment 6. Means and 95% confidence intervals are presented above. Higher numbers equal more condemnation, desire, or harm.

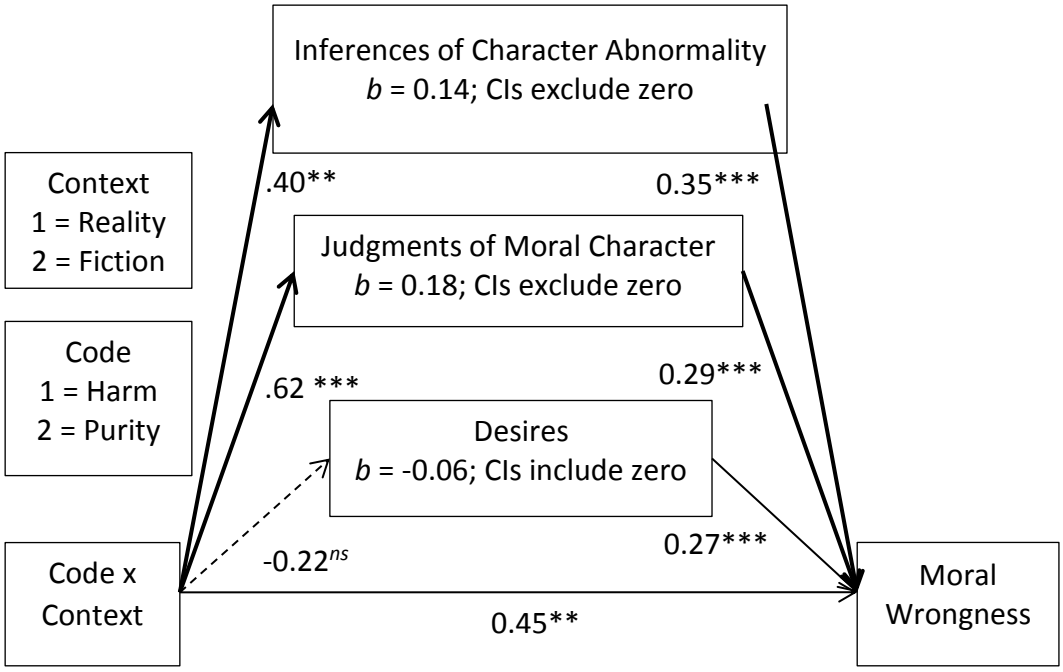


Figure 15. Unstandardized regression coefficients showing the mediating effects of moral character, abnormal character, and the presumption of desires on the interactive effect of Code x Context on moral wrongness, Experiment 5. Note: * < .05; ** < .01, *** < .001; Bold paths indicate significant indirect effects (95% CIs exclude zero).

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 by the presumption of desire (*b* = -0.06, *SE* = .04, CI = [-0.15, 0.02]). Furthermore, and as in Experiment 5, 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).

Discussion

Expanding on the findings of the former experiments, the present experiment has provided a more thorough picture of the fictive pass asymmetry effects. Firstly, these findings have generalized the results of the previous experiments by demonstrating fictive pass asymmetry effects across additional behaviors that exemplified harmful and impure behaviors.

Speaking of which, Experiment 6 addressed a potential theoretical concern about the distinctness of these two moral domains (Cameron et al., 2015). Instead of defining “harm” and “purity” code violations as the extent to which they evoked more anger or disgust, respectively, as was done in the previous experiments, they were instead defined by the extent to which participants in the pretest rated them as immoral and harmful to other people versus immoral and not harmful to other people. Experiment 6 confirmed the pretesting by demonstrating how fictive pass asymmetry effects present themselves even when the vignette selection was sensitive to this alternate perspective on moral domains.

Experiment 6’s analysis of moderated mediation replicated and expanded on the results of Experiment 5. It was indicated that the variance in the Code x Context interaction on moral wrongness is explained by the fact that fictional purity code violations, more so than fictional harm code violation, signal one as an abnormal person and an immoral person. The fictive pass asymmetry effects on moral wrongness are not, however, alternatively explained by the presumption of desires. To engage with

fictional impurities does not imply a desire to commit the act in real life any more so than engaging with fictional harm does.

Meta-Analyses of the Experiments

The last data to be reported will be the results of a series of meta-analyses of the previous experiments. These analyses will aim to do two things. Firstly, they intend to cement the reliability of the fictive pass asymmetry by demonstrating how its effects will clearly present themselves when the relevant effects sizes are aggregated across a number of the experiments from this thesis. Secondly, these meta-analyses will explore whether any specific fictional context, amongst those examined, is more, or less, responsible for the effects of the fictive pass asymmetry. This will be done by examining the specific Code x Individual Fictive Context interactions. To elaborate on this second point, each experiment in this thesis has collapsed the fictional contexts (imagination, watched in a film, performed in a video game) into a single level of fiction. This was intentional because there were no specific hypotheses of fictive pass effects between the fictional contexts. As such, the effect of fiction on moral condemnation could be more directly tested by simultaneously examining a number of popular outlets of fictional content. Nonetheless, it is worth exploring the specific Code x Individual Fictive Context interactions in order to gain a more thorough understanding of fictive pass effects and also to satiate any curiosity about the effect of any specific level of fiction. The tables that are in supplementary materials B contain these interactions across all six experiments. The sheer size of this display does, however, make it difficult to

draw any overarching conclusions, so to simplify this information the meta-analyses will systematically condense and analyze these results.

These meta-analyses were conducted using the downloadable Wilson macros for SPSS (Wilson, 2005). To prepare the datasets, the mean difference in moral wrongness, anger, and disgust of each reality vs fictive context (individual and aggregate) contrast for both harm and purity code violations was obtained. These analyses only contained the data from Experiments 3, 4, 5, and 6. Experiments 1 and 2 were purposefully excluded from the meta-analysis because they did not contain the full set of fictional contexts that was used from Experiment 3 onward.

The first meta-analysis on moral wrongness examines the mean difference in moral wrongness between real and fictional (all contexts) acts of harm and purity (Figure 16). Results support the fictive pass asymmetry hypothesis by indicating that across the included experiments, the mean difference in moral wrongness between real and fictional contexts is greater for harm violations than it is for purity violations. To interpret Figure 16, understand that the two data points represent the mean difference (with standard deviations) in moral wrongness between real and fictional acts of harm and purity. For the experiments included in the meta-analysis, the mean difference in moral wrongness between real and fictional contexts was $M = 2.83$ for harm code violations and $M = 0.88$ for purity code violations. In other words, across Experiments 3, 4, 5, and 6, there was a substantially larger gap between evaluations of real versus fictional harm than there was between real versus fictional impurities. According to Cummings and Finch (2005) a good rule of thumb is that if confidence intervals do not overlap by

more than 25%, then they are likely to be significantly different from one another. In the case of this first meta-analysis, it can be confidently claimed that there is a significantly larger reality to fiction gap for violations of harm than there is for violations of purity.

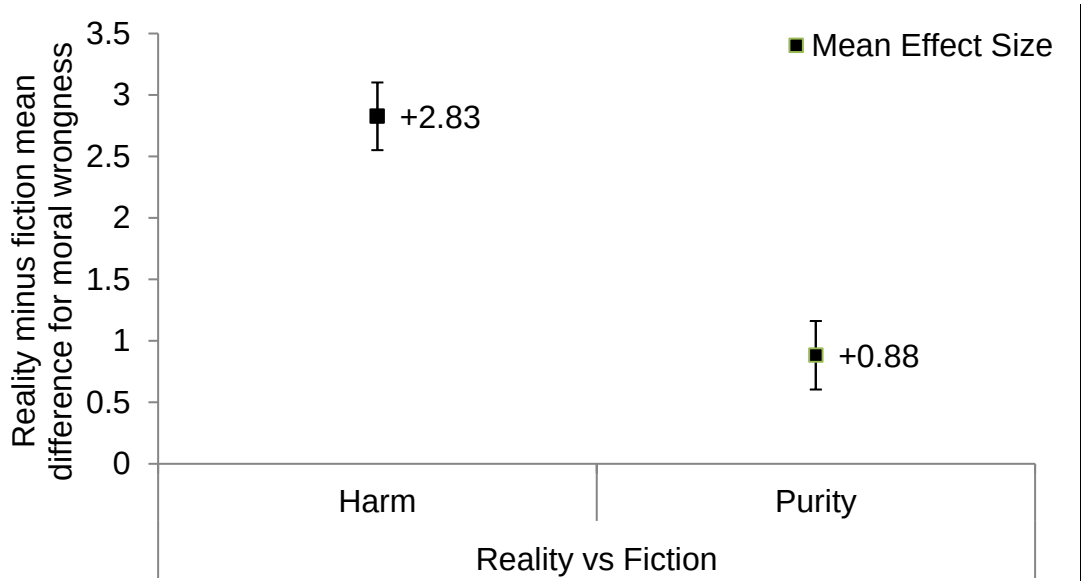


Figure 16. The mean effect sizes and 95% confidence intervals for the overall reality minus fiction contrast on moral wrongness across both moral code violations, Experiments 3, 4, 5, and 6.

The second meta-analysis examined the mean difference in moral wrongness of each reality versus individual fictional context contrast for both code violations. Looking at Figure 17 below, and again relying on Cummings and Finch's (2005) rule of thumb, it can be gathered that across the experiments, the biggest reality to fiction difference in moral wrongness was for the video game condition and the smallest reality to fiction differences was for the imagination condition. Relative to the other contrasts, the strongest fictive pass asymmetry is seen in the reality vs film contrast, but

overall, fictive pass asymmetry effects can be reliably seen in all three of the fictional contexts that were included in Experiments 3 – 6.

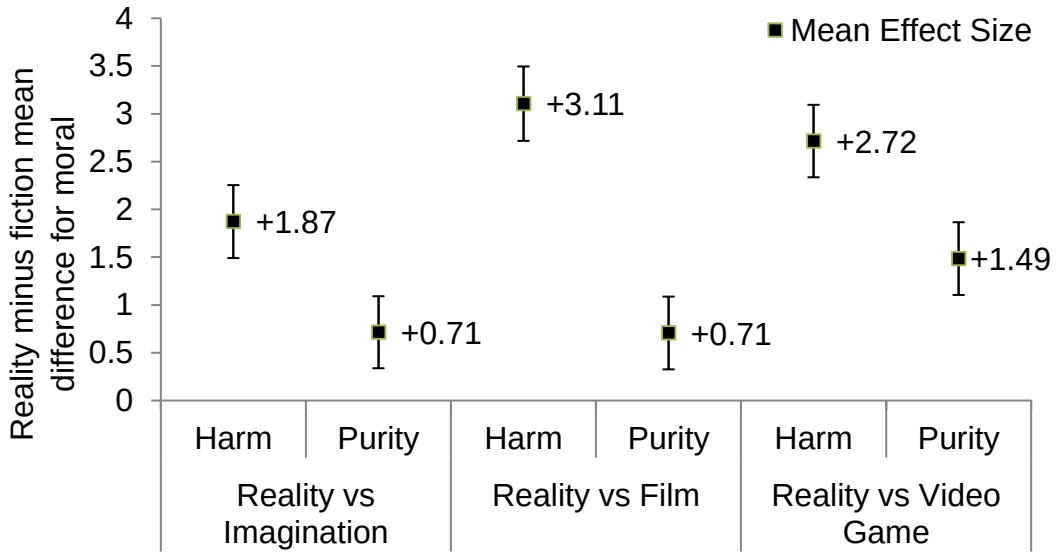


Figure 17. The mean effect size and 95% confidence intervals for each reality minus individual fictive-contrast on moral wrongness across both code violations, Experiments 3, 4, 5, and 6.

The next sets of meta-analyses focused on the emotions of anger and disgust and as before, it was conducted on Experiments 3 – 6. The data was prepared by calculating the reality minus fiction mean difference for both anger and disgust across both harm and purity code violations.

The results of these analyses show how the overall effects of the fictive pass asymmetry on anger and disgust are in line with expectations. The effects of the asymmetry are overall stronger for anger than they are for disgust. More specifically, the reality to fiction gap in disgust is about the same across both code violations, but the reality to fiction gap for anger is far greater for harm than it is for purity. The confidence intervals indicate how

the anger directed towards harm is mitigated by fiction more so than for all other effects, which are about the same as one another.

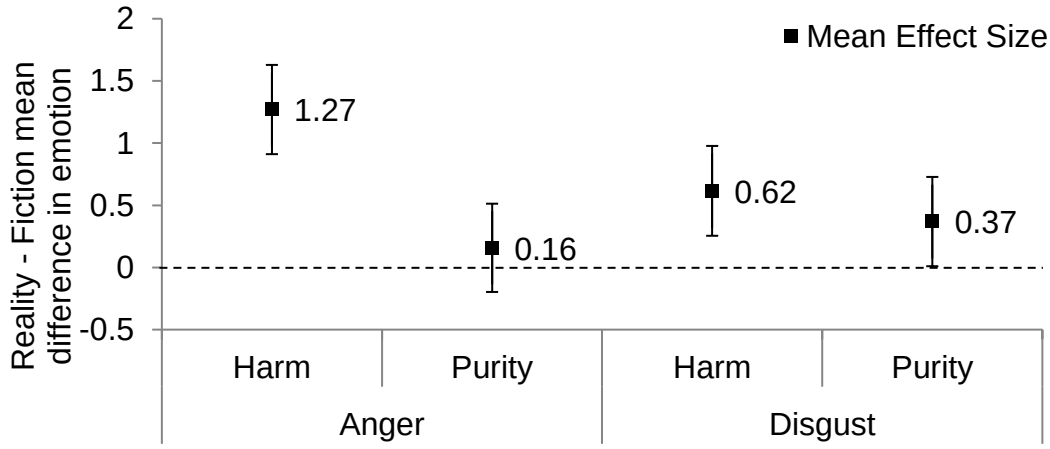


Figure 18. The mean effect size and 95% confidence intervals for each reality minus fiction contrast on anger and disgust across both code violations, Experiments 3, 4, 5, and 6.

Conclusion

This chapter has reported a final experiment and the results of a series of meta-analyses of a number of the experiments that have been reported in this thesis. Experiment 6 addressed an alternate theoretical perspective on moral codes that stood to challenge the validity of the fictive pass asymmetry effects. Cameron et al. (2015) argue that harm and so called “objectively harmless purity code violations” are confounded with one another because participants rationalize the occurrence harm in all immoral acts. This concern was initially addressed in pretesting where vignettes were selected based on the extent to which they were seen (by participants) as immoral and harmful to others versus immoral and not harmful to others. In addition to this, Experiment 6 itself explored the nature of this perceived

harm and it was found that the acts that had been previously defined as “harm code” violations were mostly strongly associated with social harms to the community or to specific others, and that these social harms concerns were predicted by anger. By contrast, the acts that had been previously defined as “purity” code violations were most strongly associated with harm to non-social entities as the self or nature and that these non-social harms were predicted by disgust. Most importantly, acts that caused social harms showed fictive pass asymmetry whereas acts that caused non-social harms did not. In effect, when the morally relevant acts were redefined as acts that caused social versus non-social harm – definitions that are in agreement with Cameron et al.’s (2015) reclassification of the moral domains – these two categories still behaved as they did when they were previously defined as violations of harm and purity.

A series of meta-analyses were also reported in this chapter that aimed to examine the effects of the fictive pass asymmetry across four of the six experiments that have been reported in this thesis. The first two analyses focused on the overall effect of the fictive pass asymmetry on moral wrongness. The first analysis followed the lead of the experiments in this thesis and examined the reality versus fiction difference (collapsed from three levels) in moral wrongness between harm and purity code violations. The second examined each reality vs individual-fictive-context contrast in order to explore whether any specific contrast was more, or less, responsible for carrying the effects of the asymmetry. Both of these analyses confirmed the overall effects of the fictive pass asymmetry and while there were some minor differences in the extent to which the individual reality versus fiction

contrasts showed effects, the asymmetry was evident in the three fictional contexts that were examined in this research: imagination, acts watched in films, and acts performed in video games.

Lastly, a meta-analysis reviewed the effects of anger and disgust on the fictive pass asymmetry. As initially predicted, effects were stronger for anger than for disgust, and the reality to fiction decline in emotion was mostly evident for anger felt towards acts of harm.

From the results of these meta-analyses, it can be concluded that harmful and impure acts are not evaluated equally in fictional contexts. When one consumes fiction that encourages the imagining of an act of harm, it is generally acceptable and not subjected to very much moral condemnation, even though the exact same acts are highly condemnable when they are actually performed. By contrast, the amount that one is condemned for using fiction that encourages the imagining of a purity code violation is relatively closer to the amount of condemnation received by one who commits an identical violation in the context of real life.

CHAPTER 7

GENERAL DISCUSSION

The primary goal of this thesis was to experimentally measure and compare moral condemnation and moral emotions towards real and fictional acts that violate moral norms of harm and purity. More generally, this research has been seeking support for the *fictive pass asymmetry*, a hypothesis which posits that evaluations of harm code violations, more so than purity code violations, will be mitigated by fictional contexts. In the strongest expression of this hypothesis, it was predicted that there would be a significant difference between the moral condemnation of real versus fictional acts of harm, but that condemnation of purity code violations would be the same across both contexts. While these ideal effects were only found in an experiment that suffered from potential floor effects, it was nonetheless consistently found that condemnation of harm was mitigated by fiction relatively more than impurities were. Although these effects were not absolute, their reliability was given further support by a series of meta-analyses.

Initially, the *fictive pass asymmetry hypothesis* was inspired by casual observations of people's reactions to media products that encourage consumers to imagine immoral behavior. It seems, at least in the western English-speaking world, that there is a greater tolerance towards the fictional depiction of violent acts than the fictional depiction of sexual acts. The research in this thesis has provided a theoretical framework and designed experiments that have offered empirical evidence for this phenomenon. More

specifically, this research has unveiled three novel findings about the nature of real versus fictional moral code violations.

First, it was reliably found across six experiments, and then confirmed in a meta-analysis of those experiments, that there is a discrepancy in moral condemnation between real and fictional acts that violate harm and purity norms. The reality-to-fiction gap in wrongness was significantly greater for harm than it was for purity. Secondly, the fictive pass asymmetry affects the moral emotions of anger and disgust to varying degrees. Anger tended to show fictive pass effects, but this was not always the case for disgust, which showed relatively more equal reality-to-fiction drops across harm and purity code violations. Lastly, the effects of the fictive pass asymmetry were partially explained because harmful fiction, unlike impure fiction, was not seen as indicating anything bad about the moral character of the one who consumes it.

This research has been most strongly influenced by literature that examines differences between the moral domains of harm and purity, the emotions of anger and disgust, and moral judgments of act and character. Previous research has examined these facets of morality individually and also established links between them (i.e. Chakroff & Young, 2015; Russell & Giner-Sorolla, 2011b; Russell & Giner-Sorolla, 2013), and from this body of work it can be gathered that condemnation of harm code violations, relative to purity code violations, is flexible and can be mitigated by situation factors. However, this body of work has only focused on acts that occur in the context of real life, and prior to the research that is presented in this thesis, it was never considered how moral judgments and emotions might cross the

line from reality to fictional contexts. The research in this thesis reinforces and expands upon the established theoretical framework that is used in moral psychology. From an applied perspective, it can reveal the consequences of imagining about, or consuming media that encourages the imagination of, harmful and impure acts.

Summary of the Experiments

The results of six experiments and a meta-analysis that are presented in this thesis have lent support to the fictive pass asymmetry hypothesis. It has been reliably found that fictional contexts such as imagination, film, and video games mitigate moral condemnation of acts that harm other people. On the other hand, fictional violations of “purity”, that only harm non-social entities, are relatively closer in condemnation to their real-life counterparts.

Experiment 1 was an initial test of the fictive pass asymmetry, and the manipulations that were used did not strictly exemplify harmful versus impure behaviors. Instead, they focused more generally on describing immoral violent acts versus immoral sexual acts in order to explore fictive pass effects in relation to behaviors as they might be typically portrayed in contemporary media. A result of this was a variety of confounding factors but nonetheless, there was initial evidence of the asymmetry as the evaluative discrepancy between fictional and real violence was significantly greater than it was for real versus fictional sex.

After this initial show of the asymmetry, Experiment 2 found similar effects by refining the manipulations to focus more precisely on evoking harm versus purity concerns and by maintaining parallelism between the

vignettes that described these two code violations. Experiment 3 lent further support to the fictive pass asymmetry with more valid manipulations. A pretest was used to select harm code violations based off the extent to which they were immoral, evoked more anger than disgust, and did not violate any bodily moral norms. By contrast, purity code violations were selected on the basis that they were immoral, evoked more disgust than anger, and were entirely consensual and free of harm. Lastly, adding imagination as a fictional context generalized the effects of the asymmetry. Experiments 4 and 5 sought explanations for the effects of the fictive pass asymmetry by measuring the extent to which fictive activity was seen as either a cue to a pre-existing bad character or as the perceived cause of future immoral acts and bad character. These two experiments identified the basic fictive pass effects on moral wrongness but also indicated that fictive pass effects can be explained because impure fiction is seen as a cue to one's bad character. In other words, one is given a "pass" for consuming harmful fiction because, unlike impure fiction, it does not suggest anything bad about the state of their moral character.

Experiments 1 through 5 repeatedly found evidence for the fictive pass asymmetry and also offered a character-based explanation for these effects. Experiment 6 was carried out in order to address an alternate perspective on the distinctness of the harm and purity domains and to test the presumption of desires as an alternate explanation of the asymmetry. Specifically, Experiment 6 aimed to address the work of Gray et al. (2012, 2014) and Cameron et al. (2015) who argue that the harm and purity violations both involve harm, and are thus confounded with one another. To

achieve this, harm and purity code violations were re-categorized as the extent to which they were seen as immoral and harming other people versus immoral and not harming other people. This is different from the categorization of the previous experiments, which identified harmful versus impure acts based on the extent to which they evoked more anger than disgust, and vice versa for purity. To validate this new categorization, the design included explicit measures of harmfulness in which participants were asked to report how much harm the behaviors in the vignettes caused to a variety of targets such as other people, the self, or nature. Experiment 6 also included measures of desire as an alternate explanation of fictive pass asymmetry effects.

Experiment 6 supported the fictive pass asymmetry hypothesis, and the inclusion of new measures of *harmfulness* suggested that the harm versus purity distinction can be alternatively characterized by the extent to which the acts harm social versus non-social entities. Effectively, it was found that “purity” violations, even those that are objectively harmless, do contain a type of harm, but it is abstract, and clearly different from the more prototypical harm that is caused by “harm” violations. The presumption of desires did not offer an alternate explanation of the asymmetry, and the results still suggest that impure fiction is more condemnable than harmful fiction because it signals information about a poor moral character.

As indicated by its name, the fictive pass *asymmetry* hypothesis predicts a difference between differences. More specifically, the difference in condemnation between real and fictional contexts should be greater for harm code violations than for purity code violations. The interaction that indicated

these effects often times took a specific form: real-life harm was - more often than not - rated more severely than real-life impurities; and in fictional contexts, impurities tended to be rated more severely, or at least equally, to harm. This pattern emerged even though pretests tried to specifically select acts that were seen as equally morally wrong in the context of real life, a finding that may indicate that harm code violations are overall more condemnable than objectively harmless violations of purity in the cultural context of the sampled population.

In the case of the experiments that were present in this thesis, the sampled population consisted of Mechanical Turk workers living in the United States. Although it has been claimed that these workers are representative of the US population (Buhrmester et al., 2010), the gender demographics across the experiments in this thesis do not necessarily reflect this. Across the six experiment there are, more often than not, more men than women in the samples. The ratio is much more reasonable, and close to an even split, in the final two experiments but it is remarkably uneven across the first four experiments, with them containing about two times more men than women. One may claim that the recruitment strategy of these early experiments favored men by advertising for an experiment about video game violence or reactions to violent events. None of the advertisements did, however, pander towards a male demographic by mentioning these stereotypically masculine themes. They were ambiguous and neutral and simply advertised for a study about understanding human behavior. Moreover, measures were taken to ensure that once an individual participated in an experiment in this series, they would be unable to participate in any future experiment.

As such, the unequal gender distribution in the early experiments is somewhat of a mystery, but it still may have contributed towards stronger fictive pass effects in the early experiments, than what would have been found with a more balanced gender profile. For instance, the main character that was being evaluated in the vignettes was typically a male and it is possible the primarily male sample of the early experiments evaluated his fictional transgressions less severely as an act of solidarity. Future experiments may control for gender and even measure the extent to which the participants themselves consume fiction that depicts immoral behavior. This would ensure that participants themselves were not being lenient towards the vignettes' main character's fictional behavior, either because of gender or because of their own video game or film habits. Even if the effects of the fictive pass asymmetry were inflated in the first experiments due to the participant profile, it should be acknowledged that these effects still strongly presented themselves in Experiments 5 and 6, where the gender distribution was much more even.

A notable improvement across these experiments was the ever-increasing gap between the emotions of anger and disgust. Anger and disgust shared a substantial amount of variance in the earlier experiments, but it was consistently reduced as the manipulations and measures became more refined. In the experiments for which there was a clear distinction between these two emotions, anger showed fictive pass asymmetry effects, but disgust – more often than not – did not. These effects, while slightly inconsistent across the experiments, were confirmed by the meta-analysis. Anger and disgust reactions differed for harm code violations in that disgust

showed a smaller drop between real and fictional contexts. In other words, amongst harm code violations anger and disgust functioned differently from one another. Anger showed a greater reality to fiction decline than disgust, meaning that when someone engaged with fictional harms, disgust tended to be the most prevalent emotional reaction. Anger towards fictional harm may diminish more than disgust does because in the absence of any actually wrongdoings, anger dissipates but disgust may remain to serve as an indicator of bad character (Giner-Sorolla & Chapman, 2017; Miller, 1997).

In spite of the above claims about the role of disgust, the experiments in this thesis did not actually include emotions as mediators between moral code violations and moral wrongness. Rather than being an oversight, however, this was an intention omission because while it is reasonable to predict that the disgust felt towards purity code violations mediates the relationship between impurities and moral wrongness, the analysis itself would be redundant.

As previously discussed, there are well-established links between purity code violations and disgust (for a review see Russell, Piazza, & Giner-Sorolla, 2013). This connection is present to the extent that disgust is an essential part of a purity code violation and one simply cannot experience purity-related concerns without feeling disgust, especially in regards to an abnormal use of one's body (Russell & Giner-Sorolla, 2013). Disgust is a core component of a purity code violation and the two constructs are simply not distinct enough to include one as a mediator of the other. In this case, to examine disgust as a mediator to impurity would equate to examining the mediating effects of a manipulation check. Of course, the inflexible nature of

disgust (Russell & Giner-Sorolla, 2011b) is instrumental in laying the theoretical groundwork for the effects of the fictive pass asymmetry. Nonetheless, so far as using mediation to help explain why purity code violations, relative to harm code violations, are not granted the same amount of moral permissibility, character is a more theoretically useful mediator. Character is more informative because it is a facet of overall moral wrongness while still being distinctly unique construct (Uhlmann, Pizarro, & Diermeier, 2015). Furthermore, character related concerns should have been evoked more strongly from purity code violations than from harm code violations (Chakroff & Young, 2015), thus allowing for a more theoretically useful model in trying to determine a key distinction between the overall moral permissibility of real and fictional moral code violations, while being a more orthogonal construct than the emotions themselves.

Theoretical and Applied Implications

The results of the experiments that are reported in this thesis have demonstrated how the contexts that surround specific norm-violating acts of harm and purity influence how people morally evaluate the acts themselves, and the individuals that commit them. The notion that fictional contexts mitigate moral condemnation is, perhaps, not a surprising concept if one subscribes to a strictly consequentialist school of thought and believes that the moral wrongness of an act is entirely contingent on its outcomes. The most noteworthy aspect of this research is, however, that the mitigation is reduced less for purity code violations and that purity code violations are related to judgments of moral character (Chakroff & Young, 2015). In

addition to this, the fourth and fifth experiments suggest that the fictive pass can be partially explained by the presumed character-related consequences of consuming immoral fiction, implying that character morality is partially responsible for downstream consequentialist concerns, even in the absence of immediate harm. Additionally, these findings have reminded us of, and reinforced, the rigid nature of purity code evaluations, relative to the more flexible nature of harm evaluations. Past literature has demonstrated the different degrees to which evaluations of these acts can be swayed by contextual factors (Gutierrez & Giner-Sorolla, 2007; Rozin, et al., 1986; Russell & Giner-Sorolla, 2011b), but the research presented in this thesis has demonstrated how these principles hold true even when the context in question is that of fiction.

These claims, however, are hindered by the fact that the sampled population consisted of people living in the Western English-speaking world. As such, it is difficult to fully address the extent to which the effects of the fictive pass asymmetry hypothesis would apply across different cultures. As an example of this, consider the differences in film rating criteria between the United States and Scandinavian countries. When determining the appropriateness of films, Scandinavian countries regulate more strictly the depiction of violent versus sexual content. The United States on the other hand regulates sexual content more strictly than violent content (Price, Palsson, & Gentile, 2014). As demonstrated in the first experiment, the general categories of violence versus sex do not perfectly translate to harm and purity violations, respectively, but this still illuminates cultural differences in the weight placed on fiction that depicts different types of morally

questionable behaviors. Further to this, is raises important questions regarding the causal nature of the fictive pass asymmetry effects. Are they innate, or do they exist and whether they exist because we have learned from the media itself that harmful acts are more justifiable than impure acts.

From a more empirical perspective, early work in cross-cultural morality put forth the idea that violations of harm are universally immoral (Turiel, Killen, & Helwig, 1987). Haidt, et al. (1993) contested the exclusive wrongness of harm by suggesting that, “the domain of morality appears to vary cross-culturally” (pg. 625). More recently, work has cemented the notion that 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 introduced by cross-cultural differences poses challenges while trying to ascertain universal truths about moral judgment. In spite of this, it is clear that the fictive pass asymmetry effects that have been found in these experiments lend strong empirical support to the casually observed discrepancy between the appropriateness of fictional harm and purity codes within the context of the sampled population. A dedicated cross-cultural study would, however, be needed to assess the true generalizability of this work.

On a similar note, it is important to consider the ecological validity of the manipulations, and the extent to which the vignettes that were used in these experiments truly depict the sorts of behaviors that are commonly portrayed in fiction. It is safe that say that one can imagine anything that is

logically possible, so virtually anything in the imagination conditions are plausible. The same, however, cannot be said of the film and video game contexts and it is possible that the participants who were assigned to these conditions needed to suspend their disbelief while reading and evaluating these acts.

The film and video game conditions may have come across as far-fetched because the vignettes in these experiments – other than those of Experiment 1 – strove for strong experimental control by specifically describing acts that violated moral norms of harm or of purity, but not of both. This experimental control may, however, have been at the cost of ecological validity considering acts that are depicted in media products rarely display acts that neatly violate a single moral code. For example, countless films and video games depict harm in the form of physical violence. Sometimes, the violence can be “sanitary” and “clean” but often times it will infringe upon the purity domain by being accompanied by blood and gore. Similarly, it is unlikely that purity code violations would appear in films and video games as they are described in these experiments, because impure bodily moral violations are frequently accompanied by an act of violence such as rape or sexual assault. In popular media products it does not, however, seem that fictional sexual content needs to be particularly abnormal or harmful to be deemed as controversial or inappropriate. Consider, again, the example of the *Grand Theft Auto: San Andreas* “Hot Coffee” modification that was described at the start of this thesis. Although the sex that was portrayed in this game was relative tame and normal, it evoked outcry as if it were much more abnormal or twisted. It is possible that even the tamest depictions of

fictional sex can be seen as taking on darker and more impure dimensions when they run the risk of falling into the hands of children. Concerns for the purity and innocence of children may conflate otherwise normal and non-controversial sexual activity with purity code violations. This is a notion that could be explored in future research by incorporating parents' views on media products with what is already understood about the fictive pass asymmetry.

In a similar vein, research on the acceptable "community standards" of fiction suggests that sexually explicit fiction that is intended to be viewed by adults is permissible so long as the content does not portray sexual violence or fetishism and it does not involve any minors (Linz, Donnerstein, Shafer, Land, McCall, & Graesser, 1995). On the one hand, this is problematic for the "Hot Coffee" modification that was used to contextualize this research, even though it is unreasonable to think that the mature rating will prevent young adults and children from playing such games. On the other hand, and more importantly, this research on community standards supports the fictive pass asymmetry by suggesting that the most condemnable fictional acts are ones that violate bodily norms, such as bondage and fetishism. More than anything, however, this reminds us that the effects of the fictive pass asymmetry can only be applied with certainty to the specific scenarios that were presented to the sampled population. Although the experiments in this thesis were high-powered and the effects were reproducible across a variety of different manipulations, one must be cautious when generalizing them too broadly. In order to begin addressing this, future research should steer away from the implausible vignettes that are commonly used by moral

psychologists and explore the fictive pass asymmetry effects by using acts that one would more realistically encounter in both real life and in fictional contexts. Secondly, a cross cultural examination of the fictive pass effects, especially between countries that place different weights on the moralization of prototypically harmful versus impure fiction, would shed light on the generalizability of the fictive pass asymmetry.

Across the experiments that are presented in this thesis, there is consistent and strong evidence for the hypothesized fictive pass asymmetry. Nonetheless, it should be acknowledged that there is a tension in regards to inherent confounds as a result of the chosen methodology by which real and fictional acts were compared. More specifically, fictional contexts in these experiments offered multiple possible targets of moral judgment and each of these targets plays a different role in the extent to which they may be responsible for the fictional transgression. In contrast, the real acts in these experiments were less ambiguous, they had a single target who played one role. Consequently, the context variable (reality vs. fiction) may be confounded with the agent's role (active perpetrator vs passive observer).

To elaborate, consider one of the vignettes that described an act occurring in the context of real-life. In this context, the target of moral condemnation was neither ambiguous nor unclear. An individual is described as committing an immoral act and subsequently, the participants were instructed to judge that person's actions. In fictional contexts, however, the target of participant's evaluations becomes less clear and there are multiple possible targets that could be evaluated: the man who is consuming the

fiction, the character within the fiction, and even the one who created the fiction (i.e. a screen writer, book author, or game designer).

Of all these possible roles involved, the present experiments only acknowledge the consumer, but even the role that the consumer played varied according to the type of fiction he was engaging with. For instance, when one imagines something one is simultaneously the creator and the consumer of the fiction, watching a film involves one passively observing another's creation, and video games strike a balance between the two for one can chose how one's character behaves, but only within the confines of another's creation.

In future work, one should consider more carefully whether or not it is appropriate to directly compare one who is passively consuming taboo media to one who is transgressing in real life. These comparisons were made in the present experiments and while distinct and reliable differences were found, it is acknowledged that these differences may not be entirely due to the context, but to the role played by the main character as an observer or as an actual perpetrator.

To shed some light on these issues, however, consider the findings of the meta-analysis of these experiments that broke the fiction variable into its respective levels (imagined, watched in a film, played in a video game). This meta-analysis sought to explore the strength of the fictive pass effects by comparing reality to each of the contexts, individually, rather than collectively as in the experiments. This meta-analysis revealed that the imagination condition showed the weakest fictive pass effects (that is, moral judgments of

imagined taboos were most close to moral judgments of real acts), the film condition showed the strongest effects, and the video game condition was in-between the two. One interpretation of this pattern, to explain why these contexts are reliably different from one another, is the role played by the agent and the extent to which the agent is responsible for generating the immoral fiction. Imagination does not equate to actual perpetration, but it certainly involves more agency than video game playing, which occurs within parameters that were set by the game's developers. Nonetheless, one may choose to act as a model citizen or as a ruthless killer in the world of a Grand Theft Auto game, and this distinguishes games from films, which are the most passive of the studied mediums. In effect, it seems that the amount of agency one has in actually generating the immoral fiction is a strong indicator of fictive pass effects.

This in turn gives rise to even more questions and avenues of future exploration that could include the role of the agent as a feature, rather than as a bug. For instance, voyeurism, as opposed to actual perpetration, could be compared with fictional contexts in an attempt to compare passive roles across different contexts. Similarly, the creator of immoral fiction could be compared with a consumer immoral fiction to compare an active vs a passive role, but within the confines of fiction. Such designs would serve two purposes. Firstly, they would control for any actual harm and allow for a more direct examination of other important factors such as the corrupting influence of fiction or the extent to which it is indicative of poor moral character. Secondly, they would allow for an even deeper understanding of fictional

taboos and allow for a wider variety of paradigm through which one can continue to explore and innovate on the fictive pass asymmetry effects.

From an applied standpoint, the research that has been presented in this thesis has shown how research in moral psychology can inform media regulation. There are various organizations such as the Motion Picture Association of America (MPAA) and the Entertainment Software Rating Board (ESRB) in North America or the British Board of Film Classification (BBFC) and Pan European Game Information (PEGI) that are responsible for systematically classifying films and video games into brackets of age appropriateness. These organizations allow consumers and parents to have an understanding of the contents of the film or video game that they are about view or purchase. The content ratings differ slightly across organizations but they all follow a similar rating scheme that ranges from *appropriate for all ages* to *adults only*. Interestingly, the rating criterion of these organizations falls in line with the effects of the fictive pass asymmetry. Products that contain fictional depictions of harm (mostly violence in the case of video games and films) are deemed appropriate for much younger ages than products that contain sexual content or other moral impurities such as drug use and gambling. Also in the most restrictive categories are acts that depict graphic harm that involve realistic blood and bodily destruction such as dismemberment and decapitation. Similar to the sexual content, these types of displays are likely to elicit concerns of disgust and purity, thus explaining their more stringent restrictions.

Interestingly, these organizations do not offer any empirical evidence for the intuitive way that they have categorized and regulated harmful versus

impure content. The experiments in this thesis can, however, justify their decision-making by reflecting public opinions about the appropriateness of fiction that depicts these two types of immoral behaviors.

In closing, the research in this thesis has demonstrated how moral code violations that occur in fictional contexts are not evaluated equally. In fictional contexts, the permissibility of fictional harms, but condemnation of fictional impurities, is undeniable. In real-life, the consequences of immoral acts are clear to see and they help to fuel the condemnation that is directed towards them. In fictional contexts though, there is a hazy ambiguity that surrounds immorality. It is “make believe”, and this makes it hard to pin down the aspects of it that evoke moral concerns. In the realm of fiction, this research has shown that fictional harm is seen as relatively benign. After all, harmful behavior in real life can be justified and the extent to which it is condemned can be modified by the context in which it occurs. It makes sense that when acts occur in fiction, they are overlooked and given a “pass” in terms of condemnation. At least, they are given a greater pass than impure behaviors which leave one with a foul taste in the mouth no matter what context, real or fictional, the act occurs in. What is more, this discrepancy can be explained by the extent to which impure behaviors signal negative information about one’s moral character. When one engages with impure fiction there are no obvious consequences and no one gets hurt, but unlike fictional harm code violations, they are still seen as wrong. This discrepancy, the fictive pass asymmetry, exists because in the absence of any consequences, character is all that is left to judge. Fictional impurities

may be make-believe, but they are “symbolically potent” (Young & Whitty, 2011 pg. 805) and will still lead to very real condemnation.

REFERENCES

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751-763. doi:10.1037/0022-3514.93.5.751

Bates, D., & Pow, H. (2013). Lanza's descent to madness and murder: Sandy hook shooter notched up 83,000 online kills including 22,000 'head shots' using violent games to train himself for his massacre. Retrieved from http://www.dailymail.co.uk/news/article-2516427/Sandy-Hook-shooter-Adam-Lanza-83k-online-kills-massacre.html

Beaupré, M., 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, PO Box 8000*

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011) Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspective on Psychological Science*, *6*(1) 3-5 doi: 10.1177/1745691610393980

Bushman, B. J., Gollwitzer, M., & Cruz, C. (2015). There is broad consensus: Media researchers agree that violent media increase aggression in children, and pediatricians and parents concur. *Psychology of Popular Media Culture, 4*(3), 200-214. doi:10.1037/ppm0000046

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

Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral

content and discrete emotions. *Personality and Social Psychology Review*. *19*(4), 371-394. doi:10.1177/1088868314566683

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

Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology, 81*(4), 697-710. doi:10.1037/0022-3514.81.4.697

Cowart, B. J. (1981). Development of taste perception in humans: Sensitivity and preference throughout the life span. *Psychological Bulletin, 90*(1), 43-73. doi:10.1037/0033-2909.90.1.43

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology, 40*, 61-149. doi:10.1016/S0065-2601(07)00002-0

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*(2), 170-180. doi:10.1037/0003-066X.60.2.170

Curtis, V., & Biran, A. (2001). Dirt, disgust, and disease: Is hygiene in our genes? *Perspectives in Biology and Medicine, 44*(1), 17-31.

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

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. doi:10.1037/0022-3514.44.1.113

Dohi, H., Yamada, A., & Entsu, S. (1991). Cattle feeding deterrents emitted from cattle feces. *Journal of Chemical Ecology, 17*(6), 1197-1203. doi:10.1007/BF01402943

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169-200. doi:10.1080/02699939208411068

Ekman, P. (1999). Facial expressions. *Handbook of Cognition and Emotion, 16*, 301-320.

Ellsworth, P. C., & Tong, E. M. (2006). What does it mean to be angry at yourself? Categories, appraisals, and the problem of language. *Emotion, 6*(4), 572-586. doi:10.1037/1528-3542.6.4.572

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgment. *Psychological Science, 22*(3), 295-299. doi:10.1177/0956797611398497

Fischer, A. H., & Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology, 93*(1), 103-115. doi:10.1037/0022-3514.93.1.103

Fischer, A., & Giner-Sorolla, R. (2016). Contempt: Derogating others while keeping calm. *Emotion Review,* 1-12. doi:10.1177/1754073915610439

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development.* Cambridge, Mass.; London: Harvard University Press.

Giner-Sorolla, R. (2012). *Judging passions: Moral emotions in persons and groups*. London, UK: Psychology Press.

Giner-Sorolla, R., & Chapman, H. A. (2017). Beyond purity: Moral disgust toward bad character. *Psychological Science*, *28*(1), 80-91. doi: 10.1177/0956797616673193

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 and Emotion, 26*(7), 1208-1222. doi:10.1080/02699931.2011.645278

Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology, 29*(56), 781-795. doi:10.1002/(SICI)1099–0992(199908/09)29:5/63.0.CO;2–3

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. doi:10.1037/a0034726

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2012). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology, 47*, 55-130.

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. doi:10.1037/a0015141

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. doi: dx.doi.org/10.1037/a0021847

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

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: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., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101-124. doi:10.1080/1047840X.2012.651387

Gray, K., & Wegner, D. M. (2010). Blaming god for our pain: Human suffering and the divine mind. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc, 14*(1), 7-16. doi:10.1177/1088868309350299

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

Gutierrez, R., Giner-Sorolla, R., & Vasiljevic, M. (2012). Just an anger synonym? Moral context influences predictors of disgust word use. *Cognition & Emotion, 26*(1), 53-64. doi:10.1080/02699931.2011.567773

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834. doi:10.1037/0033-295X.108.4.814

Haidt, J. (2003). The moral emotions. In Davidson, R. J., Scherer, K. R., & Goldsmith, H. H. (Ed.), *Handbook of affective sciences* (pp. 852-870). Oxford: Oxford University Press.

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

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. doi:10.1007/s11211-007-0034-z

Haidt, J., & Hersh, M. A. (2001). Sexual Morality: The Cultures and Emotions of Conservatives and Liberals1. *Journal of Applied Social Psychology*, *31*(1), 191-221. doi: 10.1111/j.1559-1816.2001.tb02489.x

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

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

Hart, B. L. (2011). Behavioural defences in animals against pathogens and parasites: Parallels with the pillars of medicine in humans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 366*(1583), 3406-3417. doi:10.1098/rstb.2011.0092

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

Hirsch, J. K., Webb, J. R., & Jeglic, E. L. (2012). Forgiveness as a moderator of the association between anger expression and suicidal behavior. *Mental Health, Religion & Culture, 15*(3), 279-300. doi: dx.doi.org/10.1080/13674676.2011.571666

Hogan, R. (1973). Moral conduct and moral character: A psychological perspective. *Psychological Bulletin, 79*(4), 217-232. doi:10.1037/h0033956

Horberg, E., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology, 97*(6), 963-976. doi:10.1037/a0017423

Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social–functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology, 100*(4), 719-737. doi:10.1037/a0022408

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

Kidwell, B., Farmer, A., & Hardesty, D. M. (2013). Getting liberals and conservatives to go green: Political ideology and congruent appeals. *Journal of Consumer Research, 40*(2), 350-367. doi: dx.doi.org/10.1086/670610

Kohlberg, L. (1971). From Is To Ought: How to commit the naturalistic fallacy and get away with it in the study of moral Development1. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151-235). Academic Press. doi: 10.1016/B978-0-12-498640-4.50011-1

Koleva, S., Selterman, D., Iyer, R., Ditto, P., & Graham, J. (2014). The moral compass of insecurity anxious and avoidant attachment predicts moral judgment. *Social Psychological and Personality Science, 5*(2), 185-194. doi:10.1177/1948550613490965

Kolhberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In, T. Mischel (ed.), *Psychology and Genetic Epistemology,* , 151-235. doi:10.1016/B978-0-12-498640-4.50011-1

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*(1), 151-158. doi:10.1037/0022-3514.68.1.151

Levenson, R. W., & Ekman, P. (2002). Difficulty does not account for emotion-specific heart rate changes in the directed facial action task. *Psychophysiology, 39*(3), 397-405. doi:10.1017/S0048577201393150

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 & Society Review, 29*(1), 127-168. doi:10.2307/3054056

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

Morales, A. C., & Fitzsimons, G. J. (2007). Product contagion: Changing consumer evaluations through physical contact with “disgusting” products. *Journal of Marketing Research, 44*(2), 272-283.

Nabi, R. L. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition & Emotion, 16*(5), 695-703. doi:10.1080/02699930143000437

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin, 135*(2), 303-321. doi:10.1037/a0014823

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010) Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5) 411 – 419.

Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior, 27*(2), 65-87. doi:10.1023/A:1023910408854

Park, J. H., Van Leeuwen, F., & Chochorelou, Y. (2013). Disease-avoidance processes and stigmatization: Cues of substandard health arouse heightened discomfort with physical contact. *The Journal of Social Psychology, 153*(2), 212-228. doi:10.1080/00224545.2012.721812

Piazza, J., Russell, P. S., & Sousa, P. (2013). Moral emotions and the envisaging of mitigating circumstances for wrongdoing. *Cognition & Emotion, 27*(4), 707-722. doi:10.1080/02699931.2012.736859

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

Pryor, J. B., Reeder, G. D., Yeadon, C., & Hesson-McInnis, M. (2004). A dual-process model of reactions to perceived stigma. *Journal of Personality and Social Psychology, 87*(4), 436-452.

Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91*(4), 644.

Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology, 67*(2), 206-221.

Rosenberg, S., Nelson, C., & Vivekananthan, P. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*(4), 283-294. doi:10.1037/h0026086

Royzman, E., Atanasov, P., Landy, J. F., Parks, A., & Gepty, A. (2014). CAD or MAD? anger (not disgust) as the predominant response to pathogen-free violations of the divinity code. *Emotion, 14*(5), 892-907. doi:10.1037/a0036829; 10.1037

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.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*(4), 574-586. doi:10.1037/0022-3514.76.4.574

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

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology, 67*(2), 186-205. doi:10.1037/0022-3514.67.2.186

Russell, P. S., & Giner-Sorolla, R. (2011a). Moral anger is more flexible than moral disgust. *Social Psychological and Personality Science, 2*(4), 360-364.

Russell, P. S., & Giner-Sorolla, R. (2011b). Moral anger, but not moral disgust, responds to intentionality. *Emotion, 11*(2), 233-240. doi:10.1037/a0022598

Russell, P. S., & Giner-Sorolla, R. (2011c). Social justifications for moral emotions: When reasons for disgust are less elaborated than for anger. *Emotion, 11*(3), 637-646. doi:10.1037/a0022600

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

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:10.1177/1948550612442913

Russell, P. S., & Piazza, J. (2015). Consenting to counter-normative sexual acts: Differential effects of consent on anger and disgust as a function of transgressor or consenter. *Cognition and Emotion, 29*(4), 634-653. doi:10.1080/02699931.2014.930420

Schein, C., & Gray, K. (2016). Moralization and harmification: The dyadic loop explains how the innocuous becomes harmful and wrong. *Psychological Inquiry, 27*(1), 62-65. doi: dx.doi.org/10.1080/1047840X.2016.1111121

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of*

Personality and Social Psychology, 52(6), 1061-1086. doi:10.1037/0022-3514.52.6.1061

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The 'big three' of morality (autonomy, community, divinity) and the 'big three' explanations of suffering. In A. M. Brandt, & P. Rozin (Eds.), *Morality and health* (pp. 119-169). Florence, KY, US: Taylor & Frances/Routledge.

Stevenson, R. J., Oaten, M. J., Case, T. I., Repacholi, B. M., & Wagland, P. (2010). Children's response to adult disgust elicitors: Development and acquisition. *Developmental Psychology, 46*(1), 165-177. doi:10.1037/a0016692

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58*, 345-372. doi: 10.1146/annurev.psych.56.091103.070145

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

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* (155-243). Chicago: University of Chicago Press.

Uhlmann, E. L., & Zhu, L. (2013). Acts, persons, and intuitions person-centered cues and gut reactions to harmless transgressions. *Social*

Psychological and Personality Science, 5(3), 279-285. doi:10.1177/1948550613497238

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

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

Vauclair, C., & 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

Vecina, M. L. (2014). The five moral foundations sacredness scale in men in court-mandated treatment for violently abusing their partners. *Personality and Individual Differences, 64*, 46-51. doi: 10.1016/j.paid.2014.02.021

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16*(10), 780-784. doi:10.1111/j.1467-9280.2005.01614.x

Wilson, D. B. (2005). Meta-analysis macros for SAS, SPSS, and stata. Retrieved from http://mason.gmu.edu/~dwilsonb/ma.html

Winegard, B., & Deaner, R. O. (2010). The evolutionary significance of red sox nation: Sport fandom as a byproduct of coalitional psychology. *Evolutionary Psychology, 8*(3), 432-446. doi:10.1177/147470491000800310

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*(12), 1251-1263. doi:10.1177/01461672982412001

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*, 802-820. doi:10.1177/0959354310378926

SUPPLEMENTARY MATERIALS

Supplementary Materials A

Vignettes across all codes and contexts

Reality; *Film*; **Video Game;** Imagination

Experiment 1

Sexual Violations

Mark owns a kitten and a puppy. He likes to pull off his boxer shorts and rub his genitals against his pet's fur.

Mark is watching a film about a man who owns a kitten and a puppy. The man in the film likes to pull off his boxer shorts and rub his genitals against his pet's fur. Mark enjoys watching this film.

Mark is playing a video game where he plays as a man who owns a kitten and a puppy. In the game, the character he controls likes to pull off his boxer shorts and rub his genitals against his pet's fur. Mark enjoys playing this video game.

Clark is a frotteur. This means that he goes into crowded public places like subway stations and buses and attempts to rub his erection against unsuspecting victims.

Clark is watching a film about a man who is frotteur. This means that the man goes into crowded places like subway stations and buses and attempts to rub his erection against unsuspecting victims. Clark enjoys watching this film.

Clark is playing a video game where he plays as a frotteur. This means that the character he controls goes into crowded public places like subway stations and buses and attempts to rub his erection against unsuspecting victims. Clark enjoys playing this video game.

During a large family reunion, Bruce met an attractive female cousin and introduced himself as a friend of the family. He then proceeded to seduce her and they ultimately had consensual sex. She never learned his true identity.

Bruce is watching a film about a man that meets an attractive female cousin at a large family reunion. He introduces himself as a friend of the family and then proceeds to seduce her. They ultimately have consensual sex and she never learns of his true identity. Bruce enjoys watching this film.

Bruce is playing a video game where he plays as a man who meets an attractive female cousin at a large family reunion. The character he controls introduces himself as a friend of the family and then proceeds to seduce the cousin. They ultimately have consensual sex and she never learns of his character's true identity. Bruce enjoys playing this video game.

Tom is 57 and he only finds 16 year old girls attractive. He won't sleep with anyone under 16 as he knows it's illegal but he will find girls as young as possible to have sexual relations with.

Tom is watching a film about a 57 year old man who only finds 16 year old girls attractive. The man won't sleep with anyone under 16 as he knows it's illegal but he will find girls as young as possible to have sexual relations with. Tom enjoys watching this film.

Tom is playing video game where he plays as a 57 year old man who only finds 16 year old girls attractive. The character he controls won't sleep with anyone under 16 as he knows it's illegal but he will find girls as young as possible to have sexual relations with. Tom enjoys playing this video game.

Violent Violations

Jim shot a man at point blank range with a shotgun. There was an explosion of gore as the man dropped dead.

Jim is watching a film where a man is shot at point blank range with a shotgun. There was an explosion of gore as the man dropped dead. Jim enjoys watching this movie.

Jim is playing a video game where the character he controls shoots a man at point blank range with a shotgun. There is an explosion of gore as the man drops dead. Jim enjoys playing this video game.

Marshal intentionally ran over a pedestrian at a street crossing. He heard a loud crunch and squishing sounds as the man went under the tires.

Marshal is watching a film where a man intentionally runs over a pedestrian at a street crossing. He hears a loud crunch and squishing sounds as the man goes under the tires. Marshal enjoys watching this film.

Marshal is playing a video game where the character he controls intentionally runs over a pedestrian at a street crossing. He hears a loud crunch and squishing sounds as the man goes under the tires. Marshal enjoys playing this video game.

Peter likes to attack elderly people with a device that will cause their pacemakers to malfunction.

Peter is watching a film about a man who likes to attack elderly people with a device that will cause their pacemakers to malfunction. Peter enjoys watching this film.

Peter is playing a video game where the character he controls attacks elderly people with a device that causes their pacemakers to malfunction. Peter enjoys playing this game.

Kieron is a doctor. Sometimes he enjoys to secretly euthanize people that he could just as easily treat.

Kieron is watching a film about a doctor. Sometimes the doctor enjoys to secretly euthanize people that he could just as easily treat. Kieron enjoys watching this film.

Kieron is playing a video game where he plays as a doctor. Sometimes the character he controls secretly euthanizes people that he could just as easily treat. Kieron enjoys playing this video game.

Experiment 2

Purity Vignettes

Robert is a university student who owns a piercing gun. He goes to parties and enjoys giving genital piercings to anyone who wants one.

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.

Robert plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Robert can control his character to do in this virtual environment.

In the video game, Robert controls a university student who owns a piercing gun. He controls his character to parties and enjoys giving genital piercings to anyone who wants one. Robert enjoys playing this game.

Robert likes to humiliate his girlfriend with degrading sex acts that she consents to.

Robert watches a film about a man who likes to humiliate his girlfriend with degrading sex acts that she consents to. Robert enjoys watching this film.

Robert plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Robert can control his character to do in this virtual environment.

In the video game, Robert humiliates his character's girlfriend with degrading sex acts that she consents to. Robert enjoys playing this game.

Robert, a 19 year old, works at a retirement community and has a consensual sexual relationship with a 72 year old woman.

Robert, a 19 year old, watches film about a 19 year old man who works at a retirement community and has a consensual sexual relationship with a 72 year old woman. Robert enjoys watching this film.

Robert, a 19 year old, plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Robert can control his character to do in this virtual environment.

In the video game, Robert's character works at a retirement community. Robert controls his character to have a consensual sexual relationship with a 72 year old woman. Robert enjoys playing this game.

Harm Vignettes

Robert is a university student who owns a piercing gun. He goes to parties and secretly gives small and tasteful ear piercings to people who are passed out drunk.

Robert watches a film about a university student who owns a piercing gun. The student goes to parties and secretly gives small and tasteful ear piercings to people who are passed out drunk. Robert enjoys watching this film.

Robert plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Robert can control his character to do in this virtual environment.

In the video game, Robert controls a university student who owns a piercing gun. He controls his character to parties and secretly gives small and tasteful ear piercings to people who are passed out drunk. Robert enjoys playing this game.

Robert likes to humiliate his girlfriend by saying cruel and abusive things in public.

Robert watches a film about a man who likes to humiliate his girlfriend by saying cruel and abusive things in public. Robert enjoys watching this film.

Robert plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Robert can control his character to do in this virtual environment.

In the video game, Robert humiliates his character's girlfriend by saying cruel and abusive things in public. Robert enjoys playing this game.

Robert, a 19 year old, works at a retirement community and he deceives and threatens an old woman to inherit her legacy.

Robert, a 19 year old, watches a film about a 19 year old man who works at a retirement community and deceives and threatens an old woman to inherit her legacy. Robert enjoys watching this film.

Robert, a 19 year old, plays a video game that takes place in a large and realistic environment. There are many different things, both good and bad, that Robert can control his character to do in this virtual environment.

In the video game, Robert's character works at a retirement community. Robert controls his character to deceive and threaten an old woman to inherit her legacy. Robert enjoys playing this game.

Experiments 3, 4, and 5

Purity Vignettes

Sam buys a dead chicken from the store. He masturbates with it before cooking it and eating it.

Sam buys a dead chicken from the store. He imagines masturbating with it before cooking it and eating it. He enjoys imagining this.

Sam watches a film about a man that's the same age that he is. The man in the film buys a dead chicken from the store. The man in the film masturbates with the dead chicken before cooking it and eating it. Sam enjoys watching this film.

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 buy a dead chicken from the store. He controls his character to masturbate with the dead chicken before controlling him to cook it and eat it. Sam enjoys playing this video game.

Sam works in an office with a unisex bathroom. He likes to go into the bathroom after it is used by female coworkers and lick the toilet seat.

Sam works in an office with a unisex bathroom. He likes to imagine that he goes into the bathroom after it is used by female coworkers to lick the toilet seat. He enjoys imagining this.

Sam watches a film about a man that's the same age as he is. The man in the film works in an office with a unisex bathroom. The man likes to go into the bathroom after it is used by female coworkers and lick the toilet seat. Sam enjoys watching this film.

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. The character works in an office that has a unisex bathroom. Sam likes to control his character into the bathroom after it is used by female coworkers and he controls his character to lick the toilet seat. He enjoys playing this video game.

Harm Vignettes

A woman named Sam gets into a fight with her neighbor over the removal of a tree that is on both of their properties. Sam decides to destroy her neighbor's lawn by driving her heavy truck all over it.

A woman named Sam gets into a fight with her neighbor over the removal of a tree that is on both of their properties. Sam imagines destroying her neighbor's lawn by driving her heavy truck all over it. Sam enjoys imagining this.

A woman named Sam watches a film about a woman that's the same age as she is.

The woman in the film gets into a fight with her neighbor over the removal of a tree that is on both of their properties. The woman in the film decides to destroy her neighbor's lawn by driving her heavy truck all over it. Sam enjoys watching this film.

A woman named 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 her character to do in this virtual environment.

In this video game, Sam controls a character that's the same age as she is. Her character gets into a fight with her neighbor over the removal of a tree that is on both of their properties. Sam controls her character to destroy the neighbor's lawn by driving a heavy truck all over it. Sam enjoys playing this video game.

Sam shouted at his girlfriend because she did not have enough time to put on make-up before a date.

Sam imagines shouting at his girlfriend because she did not have enough time to put on make-up before a date. Sam enjoys imagining this.

Sam watches a film about a man that's the same age as he is.The man in the film shouts at his girlfriend because she did not have enough time to put on make-up before a date. Sam enjoys watching this film.

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.

Experiment 6

Purity Vignettes

Sam has sex with a frozen chicken before cooking it and eating it for dinner. Sam enjoys doing this.

Sam imagines that he has sex with a frozen chicken before cooking it and eating it for dinner. He enjoys imagining this.

Sam watches a film about a man that's about the same age as he is. The man in the film has sex with a frozen chicken before cooking it and eating it for dinner. Same enjoys watching this film.

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 have sex with a frozen chicken before cooking it and eating it for dinner. Sam enjoys playing this video game.

Sam is a member of a cannibal society. Members of this society donate their bodies so that upon death the living members can cook and eat them. Sam enjoys being part of this club.

Sam imagines that he is a member of a cannibal society. Members of this society donate their bodies so that upon death the living members can cook and eat them. Sam enjoys imagining this.

Sam watches a film about a man that's the same age as he is. The man in the film is a member of a cannibal society. Members of this society donate their bodies so that upon death the living members can cook and eat them. Sam enjoys watching this film.

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. His character is a member of a cannibal society. Members of this society donate their bodies so that upon death the living members can cook and eat them. Sam enjoys playing this video game.

Harm Vignettes

Sam comments loudly about how fat a woman looks in her jeans. Sam enjoys doing this.

Sam imagines that he comments loudly on how fat a woman looks in her jeans. Sam enjoys imagining this.

Sam watches a film about a man that's the same age as he is. The man in the film comments loudly about how fat a woman looks in her jeans. Sam enjoys watching this film.

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 comment loudly about how fat a woman looks in her jeans. Sam enjoys playing this video game.

Sam gets into a fight with another man and he punches the other guy in the face. He enjoys doing this.

Sam imagined getting into a fight with another man and punching the other guy in the face. He enjoys imagining this.

Sam watches a film about a man that's the same age as he is. The man in the film gets into a fight with another man and punches the other guy in the face. Sam enjoys watching this film.

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. His character gets into a fight with another guy and Sam controls him to punch the other guy in the face. Sam enjoys playing this video game.

Supplementary Materials B

Code x Individual fictive context interactions across all experiments

Experiment 1

	Condition	
DV	Reality x Film	Reality x Video Game
Moral Wrongness	$F(1, 274) = 3.36$, $p = .07$, $\eta_p^2 =$	$F(1, 281) = 17.97$, $p < .001$, $\eta_p^2 = .06$
Moral Outrage	$F(1, 274) = 10.72$, $p < .001$, $\eta_p^2 =$	$F(1, 281) = 13.95$, $p < .001$, $\eta_p^2 = .05$

Experiment 2

	Condition	
DV	Reality x Film	Reality x Video Game
Moral Wrongness	$F(1, 172) = 20.16$, $p < .001$, $\eta_p^2 = .11$	$F(1, 168) = 31.97$, $p < .001$, $\eta_p^2 = .16$
Moral Outrage	$F(1, 172) = 11.72$, $p < .001$, $\eta_p^2 = .06$	$F(1, 168) = 27.74$, $p < .001$, $\eta_p^2 = .14$

Experiment 3

	Condition		
DVs	Reality x Imagination	Reality x Film	Reality x Video Game
Act	$F(1, 153) = .79$, $p = .38$, $\eta_p^2 = .005$	$F(1, 151) = .35$, $p < .01$, $\eta_p^2 = .05$	$F(1, 153) = .67$, $p = .02$, $\eta_p^2 = .04$
Character	$F(1, 153) = .02$, $p = .90$, $\eta_p^2 = .00$	$F(1, 151) = 3.79$, $p = .53$, $\eta_p^2 = .02$	$F(1, 153) = .91$, $p = .28$, $\eta_p^2 = .03$
Anger	$F(1, 153) = .03$, $p = .86$, $\eta_p^2 = .00$	$F(1, 151) = 1.31$, $p = .25$, $\eta_p^2 = .01$	$F(1, 153) = .08$, $p = .03$, $\eta_p^2 = .03$
Disgust	$F(1, 153) = 2.35$, $p = .13$, $\eta_p^2 = .02$	$F(1, 151) = .51$, $p = .48$, $\eta_p^2 = .01$	$F(1, 153) = .35$, $p = .55$, $\eta_p^2 = .01$

Experiment 4

	Condition		
DVs	Reality x Imagination	Reality x Film	Reality x Video Game
Act	$F(1, 147) = 16.51$, $p < .001$, $\eta_p^2 = .10$	$F(1, 151) = 1.32$, $p < .001$, $\eta_p^2 = .18$	$F(1, 153) = 5.08$, $p = 009$, $\eta_p^2 = .05$
Char	$F(1, 147) = 2.64$, $p = .11$, $\eta_p^2 = .02$	$F(1, 147) = 6.72$, $p = .01$, $\eta_p^2 = .04$	$F(1, 148) = .61$, $p = .44$, $\eta_p^2 = .01$
Anger	$F(1, 147) = 6.90$, $p = .01$, $\eta_p^2 = .05$	$F(1, 147) = 14.77$, $p < 001$, $\eta_p^2 = .09$	$F(1, 148) = 8.72$, $p = 004$, $\eta_p^2 = .06$
Disgust	$F(1, 147) = 1.44$, $p = .23$, $\eta_p^2 = .02$	$F(1, 147) = 4.74$, $p = .03$, $\eta_p^2 = .03$	$F(1, 148) = 3.32$, $p = .07$, $\eta_p^2 = .02$
Punish	$F(1, 147) = 14.59$, $p < .001$, $\eta_p^2 = .09$	$F(1, 147) = 19.85$, $p < .001$, $\eta_p^2 = .12$	$F(1, 147) = 8.04$, $p = .005$, $\eta_p^2 = .05$

Experiment 5

DV	Reality x Imagination	Reality x Film	Reality x Video Game
Act	$F(1, 172) = 12.16$, $p = .001$, $\eta_p^2 = .07$	$F(1,167) = 40.22$, $p < .001$, $\eta_p^2 = .19$	$F(1, 168) = 10.25$, $p = .002$, $\eta_p^2 = .06$
Char	$F(1, 173) = 2.05$, $p = .15$, $\eta_p^2 = .01$	$F(1, 168) = 5.40$, $p = .02$, $\eta_p^2 = .03$	$F(1, 169) = 5.36$, $p = .02$, $\eta_p^2 = .03$
Anger	$F(1, 173) = 7.81$, $p = .006$, $\eta_p^2 = .04$	$F(1, 168) = 11.30$, $p = .001$, $\eta_p^2 = .06$	$F(1, 169) = 3.95$, $p = .05$, $\eta_p^2 = .02$
Disgust	$F(1, 173) = 0.22$, $p = .88$, $\eta_p^2 = .01$	$F(1, 168) = 0.77$, $p = .38$, $\eta_p^2 = .01$	$F(1, 169) = 0.21$, $p = .85$, $\eta_p^2 = .001$
Punish	$F(1, 172) = 15.13$, $p < .001$, $\eta_p^2 = .08$	$F(1, 167) = 17.44$, $p < .001$, $\eta_p^2 = .10$	$F(1, 168) = 10.09$, $p = .002$, $\eta_p^2 = .06$
Conse	$F(1, 173) = 3.02$, $p = .08$, $\eta_p^2 = .02$	$F(1, 169) = 22.78$, $p < .001$, $\eta_p^2 = .12$	$F(1, 169) = 0.54$, $p = .54$, $\eta_p^2 = .01$

Experiment 6

DV	Reality x Imagination	Reality x Film	Reality x Video Game
Act	$F(1, 242) = 19.63$, $p < .001$, $\eta_p^2 = .08$	$F(1, 236) = 35.35$, $p < .001$, $\eta_p^2 = .13$	$F(1, 210) = 30.42$, $p < .001$, $\eta_p^2 = .13$
Moral Char	$F(1, 242) = 3.80$, $p = .05$, $\eta_p^2 = .02$	$F(1, 236) = 21.31$, $p < .001$, $\eta_p^2 = .08$	$F(1, 240) = 29.16$, $p < .001$, $\eta_p^2 = .11$
Abnorm Char	$F(1, 242) = 0.25$, $p = .62$, $\eta_p^2 = .001$	$F(1, 236) = 10.64$, $p = .001$, $\eta_p^2 = .04$	$F(1, 240) = 0.42$, $p = .52$, $\eta_p^2 = .002$
Desires	$F(1, 242) = 1.27$, $p = .26$, $\eta_p^2 = .005$	$F(1, 236) = 0.56$, $p = .46$, $\eta_p^2 = .002$	$F(1, 210) = 1.83$, $p = .18$, $\eta_p^2 = .01$
Social Harm	$F(1, 242) = 39.04$, $p < .001$, $\eta_p^2 = .14$	$F(1, 236) = 43.97$, $p < .001$, $\eta_p^2 = .16$	$F(1, 210) = 31.15$, $p < .001$, $\eta_p^2 = .13$
Non Social Harm	$F(1, 242) = 2.41$, $p = .12$, $\eta_p^2 = .01$	$F(1, 236) = 0.99$, $p = .32$, $\eta_p^2 = .004$	$F(1, 210) = 0.11$, $p = .74$, $\eta_p^2 = .001$
Anger	$F(1, 242) = 1.95$, $p = .16$, $\eta_p^2 = .01$	$F(1, 236) = 3.61$, $p = .06$, $\eta_p^2 = .02$	$F(1, 240) = 6.52$, $p = .01$, $\eta_p^2 = .03$
Disgust	$F(1, 242) = .006$, $p = .94$, $\eta_p^2 = .00$	$F(1, 236) = 1.21$, $p = .27$, $\eta_p^2 = .01$	$F(1, 240) = 0.23$, $p = .63$, $\eta_p^2 = .001$