UNIVERSITY OF MINNESOTA

This is to certify that I have examined this copy of a Plan B Project by

Jared D. Boots

And have found that it is complete and satisfactory in all respects, and that any and all revisions required by the final examining committee have been made.

John Blanchar, Ph.D.

Signature of Faculty Advisor

05-01-2025

Date

Disentangling Weirdness from Moral Purity:

A Test of Two Competing Theoretical Perspectives

A Plan B

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA DULUTH

BY

Jared D. Boots

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS

John Blanchar, Ph.D.

May 2025

Acknowledgements

I first want to acknowledge and thank the University of Minnesota Duluth for the opportunities the institution has provided. The financial support provided through the MAPS Graduate Student Summer Research Fellowship and the Psychology Internal Funding Research Grant have aided tremendously in my education and research. I thank Ryan Hjelle, Dr. Robert Lloyd, Dr. David Gore, Dr. Ashley Thompson, Becky Blackwood, and Dr. John Blanchar for their support during my time at UMD. Their contributions do not go unnoticed. I thank my friends and family for their continued support and for affording me the opportunity to continue my education. I thank Connie and Dennis Solvie for their past and continued generosity to myself and my family. We are forever thankful.

Abstract

Past investigations indicate that harm-based moral violations are judged to be worse when greater in magnitude or frequency, but purity-based moral violations are relatively insensitive to variations in magnitude and frequency. While some psychologists argue that this is evidence that harm and purity violations are rooted in different psychological mechanisms, others contend that all moral judgments are rooted in harm and point to confounding issues such as the association between weirdness and purity-based moral violations. The current study addresses these criticisms by experimentally varying weirdness as a feature of harm-based moral violations. A sample of 314 participants judged the wrongness of moral violations in the harm, purity, and weird-harm domains. Each violation was either low or high in magnitude, varying the frequency or amount of transgression occurring. If harm and purity moral judgments have distinct psychological mechanisms, moral evaluations of weird-harm violations low versus high in magnitude should resemble those of harm-only violations by showing sensitivity to action magnitude while also demonstrating different patterns from purity violations that instead are relatively insensitive to magnitude. In contrast, the perspective arguing for a single psychological mechanism rooted solely in perceived harm would predict that moral evaluations of weird-harm violations low versus high in magnitude should resemble those of purity violations and deviate different from harm-only violations by showing insensitivity to action magnitude. Results indicated a sensitivity to action magnitude for both weird-harm and harm-only violations but not for purity violations. These findings support a pluralistic understanding of morality suggesting the use of distinct psychological mechanisms in moral judgment dependent upon the type of moral transgression that has occurred.

Keywords: Moral cognition, purity, harm, weirdness, moral monism, moral pluralism

Table of Contents

List of Tables

List of Figures

List of Appendices

Addressing the Moral Purity-Weirdness Problem: Testing Competing Predictions from Models of Moral Monism and Pluralism

Psychologists debate whether moral judgments rely on different psychological mechanisms for harm versus purity violations or whether all moral judgments rely solely on the amount of harm done (Rottman & Young, 2019; Schein & Gray, 2017). Whereas purity violations often involve taboo norm violations of sexual, disgusting, or religious content that speak to contamination (Graham et al., 2009; Haidt et al., 1993), harm violations involve traditional acts of physical or psychological harm (Schein & Gray, 2017). Interestingly, moral judgments of harm and purity violations yield different patterns when dosage is varied in size. Moral condemnation typically fluctuates based on transgression severity, with harmful acts judged worse when more frequent or larger in magnitude. Impure acts, in contrast, are strongly condemned irrespective of their frequency and magnitude (Rottman & Young, 2019). Some scholars suggest this is evidence of a psychological mechanism unique to judgments of purity violations (Graham et al., 2013; Uhlmann & Zhu, 2013) but others point to confounding factors and instead argue that moral judgment rely on the amount of perceived harm regardless of the type of violation (Gray et al., 2023; Schein & Gray, 2017). In the present study, I test competing predictions from these two perspectives to determine whether judgments of harm- and purity-based moral violations rely on a common singular psychological mechanism or different mechanisms.

Is Morality About More Than Just Harm?

Although there is widespread agreement that morality involves considerations of harm, several psychological frameworks argue for the existence of other, independent moral domains. Purity is one prominent alternative domain that favors the avoidance of disgust and acquisition of

sanctity (Haidt et al., 1993). Violations in this domain often involve no (obvious) physical harm (e.g., "A man masturbates with a dead chicken carcass"), no psychological harm (e.g., "Two siblings engaging in consensual sex"), and uncommon taboo situations (e.g., "A family eating their pet dog that passed of natural causes;" see Clifford et al., 2015; Haidt et al., 1993; Rottman & Young, 2019). Purity morality most closely aligns with the binding moral foundation of sanctity/degradation (Haidt, 2012; Haidt & Graham, 2007). Harm violations, on the other hand, involve obvious physical or psychological harm (e.g., "A girl laughs when she realizes her friend's dad is a janitor," "A person throws a rock at a farm animal;" see Clifford et al., 2015; Rottman & Young, 2019). These violations tend to produce injustice and/or threaten the desire to protect people, which align with the individualizing moral dimensions of fairness/cheating and care/harm, respectively (Haidt, 2012; Haidt & Graham, 2007). These distinctions between harm and purity violations suggest the possibility of a pluralistic psychological framework of morality that extends beyond concerns about harm.

Theories of Moral Monism

The rationalist tradition. A rationalist perspective of moral judgment was emphasized by Swiss psychologist Jean Piaget (1932), who believed morality develops as children advance their ability to think and construct more complex cognitive models of the world. Kohlberg (1971) agreed with Piaget's approach, focusing not on the outcomes of people's judgments but rather how they reasoned toward those judgments. Essentially, Kohlberg argued for a monistic theory in the sense that moral cognition is simplified to justice *reasoning* and proposed that moral judgment develops in three levels: preconventional, conventional, and postconventional. Whereas the preconventional level concerns self-interest based on achieving rewards and avoiding punishments (e.g., "Stealing is wrong because I will get in trouble"), the conventional

level advances to applying considerations of social norms and the maintenance of law and order into moral reasoning (e.g., “Stealing is wrong because it is illegal”). The postconventional level, in contrast, concerns with the internal principles of what is just, based on the perception of law and social norms (e.g., “Stealing is wrong because it goes against my values”). At all stages, the single cognitive process of justice reasoning drives moral reasoning, and as people mature and gain higher-order cognitive skills, such as perspective-taking, their moral judgments become more definitive (Smith, 2018).

Theory of Dyadic Morality. More recently, moral monism has been championed by the Theory of Dyadic Morality. It states that moral judgments rely on a cognitive template of harm and the perception of the amount of harm in situations of, “An intentional agent causing harm to a vulnerable patient” (Schein & Gray, 2017). In other words, when a person asserts that an action is morally wrong, he or she is making a claim that harm has occurred. Though monistic in the sense that harm is perceived to be at the center of moral judgment, some describe the theory as pluralistic due to variation in the values and perceptions of harm amongst individuals. Importantly, the Theory of Dyadic Morality avoids weird or uncommon situations that typically occur with purity violations through the justification of perceived harm being the only factor that ought to be considered (Gray et al., 2023; Schein & Gray, 2017). However, other researchers disagree and point to research highlighting different behaviors of harm and purity violations in the face of moral judgments (Graham et al., 2013; Graham et al., 2018; Rottman & Young, 2019).

Theories of Moral Pluralism

The affect revolution. As an alternative to rationalist theories of morality, psychologists began to recognize the importance of affect in the 1980s. For instance, Haidt and colleagues

(1993) investigated how people react to offensive but victimless transgressions and noted that in some cultures affective reactions to transgressions better explained the moral judgment as opposed to the perceived harmfulness. As a result, Haidt and colleagues proposed that affect should be recognized as a source of moral judgment because of its influence on the decision process. Later, Haidt made note of "affective precursors" for judgments relating to morality, one example being distaste as an affective precursor to disgust (Haidt, 2003). This connection is further emphasized by the feeling of disgust associated with some moral judgments (Rozin et al., 2008). Additionally, the importance of affect in moral judgment has been supported in Greene and Haidt's (2002) work utilizing neuroimaging methodology to link moral judgment to amygdala activation, an area of the brain associated with affective processing.

Social intuitionism. The social intuitionist model of moral judgment states that moral judgments result from quick moral intuitions (or "gut feelings") followed retroactively by slow and deliberative moral reasoning (Haidt, 2001). This model is derived from Hume's (1777) work arguing that moral judgments result from the sentiments that individuals hold, not the knowledge they have acquired through reason and debate. For this reason, emotion becomes a larger factor in moral judgment and opposes the rationalist perspective. Social intuitionism can thus explain the irrational responses to purity dilemmas such as people most often refusing to lick a thoroughly cleaned flyswatter (Olatunji et al., 2008). Haidt (2001) compared moral judgment to an "Emotional Dog and its Rational Tail" as a metaphor for his theoretical model whereby the fast emotional appraisal of situations is followed by slow deliberative reasoning like a dog that wags its tail when happy. Haidt and colleagues (2000) administered a series of purity dilemmas to participants whose responses largely followed this pattern, and counterarguments by a research assistant undermining participants' supposed reasoning were unable to sway their

judgments. Hence, they coined the term the "Moral Dumbfounding Effect" — when people's rationale for their moral judgment is directly challenged, they fail to abandon the judgment and express dumbfoundedness where their reasoning is ineffable (see also Haidt, 2012). Difficulty explaining moral convictions based in the purity domain, as opposed to the easily explainable convictions based in the harm domain, is consistent with a pluralistic perspective on morality.

Moral Foundations Theory. Alongside the views that affect is involved in moral judgment and that judgments are in some cases irrational, pluralists point to individual moral categories that dictate moral judgments. One of the broadest pluralist theories regarding these categories is the Moral Foundations Theory that lays out five unique "foundations" that form the basis for moral judgments (Graham et al., 2013). The theory embraces the social intuitionist model's perspective that immediate moral intuitions are followed by retroactive reasoning (Graham et al., 2013; Haidt, 2001). Additionally, Moral Foundations Theory uses both culture and social experiences to explain the ways that these intuitions present themselves as well as how the individual foundations benefit the prosocial goals of the individual.

The five foundations consist of values that are used to ascribe a domain or type to a moral dilemma and explains the adaptive challenge at hand. The care/harm foundation for instance is thought to help individuals protect their children and themselves from physical and emotional harm and is accompanied by the virtues caring and kindness (Graham et al., 2013). This foundation is "activated" when judging situations in which physical or mental harm is a threat to an individual and weighs situational aspects such as intent, quantity, and magnitude of the threat (Cushman, 2008). The foundations fairness/cheating, loyalty/betrayal, and authority/subversion are all believed to have implications for social relationships (Graham et al., 2013). The

sanctity/degradation foundation, involved in moral purity, plays a less clear role considering the perspective of the moral dumbfounding effect (Haidt et al., 2000).

Sacred and protected values. Other models highlight notions of sacredness and how they yield to protected values which are used in moral judgment (Tetlock, 2000). When considering the values that an individual holds sacred it becomes clear that considerations beyond harm are present. Sacred values are the beliefs or standards that an individual lives by that extend beyond the materialistic world such as one's commitment to love, honor, or justice (Tetlock, 2003). Secular values on the other hand are beliefs that reside in the convenience of life and may take the form of money or other worldly pleasure. Tetlock (2003) presents the Sacred Values Protection Model to help distinguish the two, positing that when sacred values are threatened by secular values, people will struggle to protect themselves from the impure thoughts associated with the threat. This relationship was demonstrated in a study presenting participants with scenarios of common sacred values (i.e. upholding religious beliefs, code of honor, etc.) then asking them to answer questions protecting the sacred value at the cost of a secular value (i.e. financial gain, compensation, etc.) (Baron & Spranca, 1997). Participants often insisted on upholding the sacred value though struggled with consequential loss of the secular value. These findings are further emphasized when threats are presented as tradeoffs, suggesting that there are different psychological mechanisms at play depending upon the details of a situation (Baron & Spranca, 1997).

Distinguishing Purity from Harm

Pluralistic moral theorists contend that judgments of harm and purity violations appear to be the products of distinct psychological mechanisms. Evidence presented in support of this claim involves intuitively harmless wrongs, feelings of disgust, and inconsistent relationships

with action severity. That is, pluralists point out that many purity violations can be categorized as harmless wrongs (Haidt et al., 1993), are often connected to or driven by feelings of disgust (Horberg et al., 2009; Rozin et al., 2008), and evoke distinct patterns that deviate from those of harmful violations when important features like intent and severity are varied (Rottman & Young, 2019; Young & Saxe, 2011).

Harmless Wrongs

Harmless wrongs are often described as acts that skew from prosocial behavior but do not disadvantage others (Fitouchi et al., 2023). These harmless wrongs are most representative of purity violations and vary widely, ranging from the consumption of drugs and alcohol to masturbating with a chicken carcass (Rottman & Young, 2019). Although moral monists argue that these activities *can* be harmful, for instance, if consensual incest *does* lead to malformed offspring (Gray et al., 2023), another explanation for why people condemn these acts is that engaging in them may be an indication of other antisocial behaviors that put those around the perpetrator at risk (Fitouchi et al., 2023). Another explanation may be that people simply find acts such as consensual incest disgusting which ultimately leads to the moral dumbfounding effect (Haidt et al., 2000).

Feelings of Disgust

When making purity-based moral judgments, ample research points to disgust as a common component that is not present in the judgment of harm-based transgression (Graham et al, 2013; Haidt et al, 1997; Wagemans et al, 2018). Rozin and colleagues (2008) extend that the sanctity/degradation component of Moral Foundations Theory, which is believed to be the source of purity-based moral judgments, is also correlated with disgust. While the sanctity/degradation foundation aids in distancing oneself from unclean, tainted, or diseased foods or items, moral

disgust manifests as an incentive to distance oneself from situations such as those involving disturbing sexual acts, gore, or mistreatment of carcasses (Haidt et al., 1997; Rozin et al., 2008). Feelings of disgust for situations that are not threatening to the physical self puzzles researchers as it does not serve as a "behavioral immune system" in the same way that reactions to unclean or contaminated items do (Rozin et al., 2008; Schaller & Park, 2011).

Insensitivity to Severity

Dosage. Another hallmark of purity-based moral judgments is their relative insensitivity to violation dosage (i.e., action magnitude or frequency) when compared to harm-based moral judgments. Rottman and Young (2019) demonstrated this by presenting participants with harm- and purity-based moral violations and asking them to rate their moral wrongness. They randomly varied the dosage level of each violation by either its frequency (e.g., sexual intercourse with a goat once vs. frequently) or magnitude (e.g., eating a small vs. large amount of flesh from dead person). The results indicated a relative insensitivity to low versus high dosage levels for purity violations, while harm violations exhibited significantly higher immorality ratings for high dosage compared to low dosage violations (Rottman & Young, 2019). Additional experiments employing dosage manipulations with a smaller difference between high and low dosage levels replicated these findings that immorality ratings of purity violations were relatively more stable across dosage levels (Blanchar et al., 2025; Rottman & Young, 2019). Consequently, Rottman and Young (2019) proposed the mere-trace hypothesis wherein moral judgments of purity violations rely solely on whether any amount of impurity is present (Rottman & Young, 2019).

Intent. Similar to dosage, intent appears to impose less influence on purity-based moral judgments than harm-based moral judgments. Several studies indicate that purity violations are insensitive to intent when morality is considered (Sweetman & Newman, 2020; Young & Saxe,

2011). Young and Saxe presented harm- and purity-based moral violations with varying levels of intent and outcome. Examples such as intentionally or accidently giving someone with a peanut allergy food with peanuts were used for the harm-based moral violations, while intentionally or accidently participating in consensual sex with a sibling was an example used for the purity-based moral violations. Results indicated that, relative to purity violation, harm violations receive significantly lower ratings of moral condemnation when they occurred accidentally compared to intentionally (Young & Saxe, 2011). Both intentional and accidental purity-based moral violations were rated similarly immoral. Additionally, the outcome of these purity-based moral violations appeared to be the driving force for moral judgment. That is, intentional attempts to commit a purity violation with only the perception of a successful attempt by the perpetrator were not rated as immoral as an unintentional attempt that was successful (Young & Saxe, 2011). These findings offer support for the mere-trace hypothesis in that moral judgments of purity violations were strongly reliant on whether any purity transgression occurred (Rottman & Young, 2019).

Mere Contemplation. Is contemplating an immoral action comparable to performing it? In Tetlock's (2003) study comparing how people react when they must choose between sacred and secular values, stark differences emerge. Tetlock (2003) observed that when people are presented with taboo trade-offs, those that pit sacred and secular values against one another, they would feel more impure the longer they spent contemplating the trade-off. For tragic trade-offs, those that pit one sacred value against another sacred value, the less time people spent deliberating over their decision, the more impure they would feel, despite no possibility to appease both protected values. This seems to indicate that when you can uphold a sacred value over a secular one, a quick decision is the only way to escape feelings of impurity (Tetlock,

2003). Interestingly, failing to deliberate between two sacred values for long enough leads to feelings of impurity despite no logical benefit to lengthened deliberation. Even when the factual evidence of no additional benefit to longer deliberation is presented, the feeling of impurity still resides, suggesting again, an emotional source to purity (Haidt et al., 2000).

Confronting the Weirdness-Purity Problem

Ample research has presented evidence that moral purity violations elicit unique response patterns when compared to those involving harmful violations (Blanchar et al., 2024; Rottman & Young, 2019). This work has received numerous criticisms, especially regarding the mere-trace hypothesis. For instance, Schein and Gray (2017) believe that moral condemnation does not involve whether a trace of impurity is involved, rather it is the result of perceived harm, norm violations, and negative affect. Additionally, analysis of the questions used in the findings that support purity's insensitivity to dosage and intent have been deemed a confound as the structure and content of the researcher created questions may be driving the unique behavior that purity exhibits (Gray & Keeney, 2015). This research indicated that the foundations that the Moral Foundations Theory scenarios were aiming to represent were not accurate and that "impure" is a synonym for "morally wrong." Other research indicates disgust as a possible confound in the findings regarding purity. One study indicated that disgust sensitivity is linked to purity which falls into the sanctity/degradation foundation of Moral Foundations Theory, though it is not strongly linked to any of the other foundations (Wagemans et al., 2018). Another focal criticism points out that purity-based violations are weird compared to the harm-based violations, which may drive the insensitivity to dosage and intent variations (Schein & Gray, 2017).

Platypus Moral Psychology

First, “weirdness” in the context of moral judgment research is defined as, “the extent to which an act is weird, bizarre, or unusual” (Gray & Keeney, 2015). Considering both the structure and weird aspects of researcher created questions used in research on both harm- and purity-based morality, strong opinions emerge regarding the behavior and differences between the two domains. Those in support of the Theory of Dyadic Morality suggest that inherent weirdness likely explains why purity violations are insensitive to dosage and intent– dubbed the weirdness overlap hypothesis (Schein & Gray, 2017; Wagemans et al., 2019). This hypothesis argues that both the disgust and purity related scenarios of the Moral Foundations Questionnaire may be weirder than those of other domains. As a result, these scholars claim that so-called purity transgressions ought to be left outside the scope of morality research as a whole stating that, “Building a model of moral judgment based upon bizarre acts such as consensual incest and chicken-sex is like building a model of mammals based upon the platypus. These bizarre scenarios are certainly interesting, but—like platy- puses—are so precisely because they are unrepresentative of their broader category,” and that while focusing on harmful acts, “We should resist a ‘platypus moral psychology’” (Schein & Gray, 2017).

Weirdness as a Confound

Considering the weirdness overlap hypothesis, it is important to flesh out whether weirdness is responsible for the unique response patterns exhibited from purity-based moral violations. Weirdness is implicated to be a large factor in disgust and purity scenarios, though it is not clear whether it is the sole factor that distinguishes purity from the other moral foundations (Wagemans et al., 2019). Considering the differences between the harm and purity violations of the Moral Foundations Theory, the answer to the debate may be held in violations that are both weird *and* harmful since removing weirdness from purity violations is challenging.

Weird but Harmful: Competing Theoretical Predictions

Collecting data on the perceived morality of harm, purity, and weird-harm moral violations could reveal support in one of two directions. The first possibility is support for a moral monistic perspective which claims that all moral judgments are derived from perceived harm and weirdness is a merely a confound of an illusory class of so-called "purity" violations. In this case it is expected that the weird-harm violations will exhibit response patterns that resemble those of the purity violations, suggesting that weirdness is indeed the driving force for the insensitivity to other factors. The second possibility is support for a moral pluralistic perspective which contends that impurity is its own unique foundation or type of moral judgment. In this case it is expected that the weird-harm violations will exhibit response patterns that resemble harm-only violations, indicating that weirdness cannot explain why judgments of purity violations are insensitive to dosage and that impurity may in fact represent a distinct psychological foundation.

Present Study

To address the debate between moral monism and pluralism, the present study aims to modify the methodology of Rottman and Young's (2019) study to investigate the distinction between judgments of harm- and purity-based violations that vary in dosage. Moral violations were sourced from the Moral Foundations Vignettes and previous studies using similar methodology then were modified to be weirder and reduce the possible confounds previously mentioned (Clifford et al., 2015; Graham et al., 2009; Gray & Keeney, 2015; Rottman & Young, 2019). A pilot study was conducted to select stimulus materials for use as harm-only, purity, and weird-harm moral violations (for details, see "Pilot Testing" below). Weird-harm violations included the traditional harm violations that were modified to be weird. For example, "A man

throws one microwave at a cow" includes a harmful act (throwing a physical object) that also happens to be odd or bizarre (the object thrown is a microwave). Using weird-harm violations in addition to harm-only violations allows me to determine whether weirdness influences moral condemnation as previously speculated for the purity violations (Wagemans et al., 2019). Following Rottman and Young (2019), all violations were experimentally varied by low and high dosage to test whether adding weirdness to harm violations yields an insensitivity to dosage, similar to that of purity violations in previous studies, or whether weirdness bestows no effect on harm, indicating that impurity and harm operate using distinct psychological mechanisms (see Figures 1 and 2). I pit two competing predictions against one another to see which is best supported. The first predicts weird-harm violations will exhibit sensitivity to dosage when rated on immorality, akin to people's responses to the harm-only violations observed by Rottman and Young (2019). The second predicts that the weird-harm violations will exhibit insensitivity to dosage like people's responses to purity violations, indicating that the weirdness of the purity violations is a confound that explains the supposed mere-trace phenomenon (Schein & Gray, 2017).

Method

Preregistration

The study design, hypothesis, sample size, exclusion criteria, and data analysis plan were preregistered at AsPredicted.org: https://aspredicted.org/2g4g-8kxm.pdf (see Appendix A).

Participants

A target sample of 330 participants was recruited for a study investigating the perceived moral wrongness of moral violations via Prolific's online platform in exchange for $1.20. Sample size was determined based on an *a priori* power analysis using the *simr* package (Green

& MacLeod, 2016) in R (R Core Team, 2023). This sample size can detect a modest two-way interaction effect ($d = 0.20$) in a linear mixed-effects model matching my experimental design with 99.1% power based on 1,000 simulations (see Figure 3). After preregistered exclusions, a final sample of 314 participants was obtained for analyses.

Data collection began following approval from the University of Minnesota Institutional Review Board and in line with the ethical guidelines of the American Psychological Association. Inclusion criteria consisted of participants being ages 18 years or older, a native English speaker, and living within the United States at the time of participation. Additionally, following Rottman and Young (2019), participants were excluded if they reported English as a second language or failed either of two attention check items ("A person destroys the entire planet" and "A person gives money to a charitable organization;" for details, see Appendix I).

Design and Procedure

The study was administered online using Qualtrics survey software. Upon providing informed consent, participants completed a questionnaire that included basic demographic questions regarding age, race/ethnicity, gender, and English as a primary or second language. Next, participants indicated the wrongness ("How morally wrong was this action?") of 10 harm (e.g., "A person punches [one person/20 people]"), 10 purity (e.g., "A person inappropriately touches [a corpse/20 corpses]"), and 10 harmful but weird moral violations (e.g., "A person captures[s] [a single/20] hungry bear[s] from a bicycle riding circus act and releases [it/them] inside an elementary school.") using 1 (*Not At All*) to 100 (*Extremely*) response scales. Each moral violation was experimentally varied by dosage (low vs. high) regarding its magnitude (e.g., "One gram" vs. "A quarter pound") or frequency (e.g., "one person" vs. "20 people"). Therefore, this experiment utilized a 3 (Domain: Harm, Purity, Weird-Harm) × 2 (Dosage: Low,

High) factorial design all within-subjects.[1] Ten violations from each domain were presented to participants, each randomly varied as low or high dosage versions.

Stimulus Materials

Harm and purity moral violations were sourced from previous studies (Clifford et al., 2015; Graham et al., 2009; Gray & Keeney, 2015; Rottman & Young, 2019). Weird-harm moral violations consisted of harm violations modified to be weird, odd, and bizarre. All violations were chosen based on the results of a pilot study designed to discern the degree of perceived harm, impurity, and weirdness involved in the moral violations. Additionally, following Rottman and Young (2019), two attention check items were included amongst the moral violations: "A person destroys the entire planet" and "A person gives money to a charitable organization" (1 = *not at all [morally wrong]*, 100 = *extremely [morally wrong]*). Those that failed either attention check by responding with less than 50 to the former or higher than 49 for the latter were excluded from all analyses. All stimulus materials are available in Appendix H.

Pilot Testing

A pilot study was conducted to ensure the validity of each type of moral violation. A sample of 176 participants were recruited from Prolific with the same inclusion criteria applied from the focal experiment. Participants were presented with 40 moral violations in random order (10 harm, 10 purity, and 20 weird-harm) and compensated with $2.00 for completing the 10-minute survey. Specifically, they rated each moral violation for its perceived harmfulness ("How harmful is this action?"), impurity ("How impure is this action?"), and weirdness ("How weird (i.e. unusual, bizarre, odd) is this action?"), using 1 (*Not At All*) to 7 (*Extremely*) response scales.

[1] Including weirdness as a separate factor in a 2 (Domain: Harm, Purity) × 2 (Dosage: Low, High) × 2 (Weirdness: Low, High) factorial design was infeasible due to criticisms of the purity domain having an inherent and irremovable aspect of weirdness (Schein & Gray, 2017).

Whereas items assessing harmfulness and impurity were borrowed from Rottman and Young (2019), items assessing weirdness were adapted from Wagemans et al., (2019). I identified 10 moral violations representing each of the harm, purity, and weird-harm categories (totaling to 30 scenarios) for use in the primary experiment. Harm violations were selected based on high ratings of perceived harm and low ratings of impurity and weirdness, purity violations were selected based on high impurity and low harm ratings, and weird harm violations were selected based on high perceived harm and weirdness ratings and low impurity ratings. Available in Tables 1 and 2, linear mixed effects models with simple effects comparisons supported these selections decisions. Harm-only and weird-harm items were rated as similarly harmful and impure but significantly more harmful and less impure compared to the purity items. Importantly, weird-harm and purity items were rated as similarly weird but significantly weirder compared to the harm-only items.

Analytic Strategy

Data were inspected and filtered for cases that passed my pre-registered exclusion criteria, including whether participants that failed either of the two attention checks. All analyses were computed in R version 4.3.1 (R Core Team, 2023). I fitted a linear mixed-effects model using the *lme4* (Bates et al., 2015) and *lmerTest* packages (Kuznetsova et al., 2017). As recommended by Barr et al., (2013), I began with the maximal random-effects structure that includes random intercepts and slopes of participants and items; the random effects structure was reduced in complexity as needed to allow the model to converge and avoid singular fit issues. Bonferroni corrections were applied to simple effects tests to control for rising family-wise error rate via the *emmeans* package (Lenth, 2023). Effect sizes (*d*) were calculated based on Westfall

et al. (2014). Assumptions of linearity, multivariate normality, and homoscedasticity were evaluated with the *performance* package (Lüdecke et al., 2021).

Results

Following my pre-registered analysis plan, a linear mixed effects model was fitted with moral domain (harm, purity, weird-harm), dosage (low, high), and their interaction as fixed-effects predictors of wrongness ratings.[2] The model included random intercepts of participants and items. The model revealed significant main effects of domain, $F(2, 9092) = 18.51$, $p < .001$, and dosage, $F(1, 9126) = 194.61$, $p < .001$. However, both were qualified by a significant domain by dosage interaction, $F(2, 9130) = 43.04$, $p < .001$ (see figure 4).

Simple effects tests with Bonferroni correction compared moral judgments of low and high dosage violations within the harm, purity, and weird-harm domains, respectively. I observed significant differences between low and high dosage levels for harm violations (*M*s = 68.45 vs. 80.19, *SD*s = 29.65 vs. 24.38), $b = 11.109$, $SE = 0.815$, $t(9132) = 13.63$, $p < .001$, 95% CI for b [9.51, 12.71], $d = 0.39$, and weird-harm violations (*M*s = 72.73 vs. 81.88, *SD*s = 29.32 vs 24.19), $b = 7.873$, $SE = 0.814$, $t(9132) = 9.68$, $p < .001$, 95% CI [6.28, 9.47], $d = 0.27$. That is, high dosage led to harsher moral judgments compared to low dosage for both harm-only and weird-harm violations. Conversely, no difference emerged between low ($M = 73.37$, $SD = 31.61$) and high dosage levels ($M = 75.27$, $SD = 31.87$) for purity violations, $b = 0.67$, $SE = 0.814$, $t(9132) = 0.82$, $p = .412$, 95% CI [-0.93, 2.26], $d = 0.02$. As illustrated in Figure 4, these data support the differing psychological mechanisms perspective as opposed to the weirdness as a confound and harm dependent perspective (refer to competing predictions depicted in Figure 1 and 2).

[2] See Table 3 for descriptive statistics.

Lastly, my pre-registered analysis plan called for custom contrasts with Bonferroni correction to test whether differences in moral judgments across dosage level (low vs. high) vary by moral domain (Harm: high-low vs. Purity: high-low, Harm: high-low vs. Weird-Harm: high-low, and Purity: high-low vs. Weird-Harm: high-low). First, the differences across dosage levels for harm violations were similar to that of weird-harm violations, albeit significantly different, $b = 3.24$, $SE = 1.15$, $z = 2.809$, $p = .015$, 95% CI [0.48, 5.99], $d = 0.11$. Most importantly, differences across dosage levels for harm violations were significantly larger than those of purity violations, $b = 10.44$, $SE = 1.15$, $z = 9.062$, $p < .001$, 95% CI [7.68, 13.20], $d = 0.36$, and differences across dosage levels for purity violations were significantly smaller than those of weird-harm violations, $b = -7.21$, $SE = 1.15$, $z = -6.257$, $p < .001$, 95% CI [-9.96, -4.45], $d = -0.25$. In other words, harm-only and weird harm violations were both rated as *more* morally wrong when the frequency or magnitude of the violation was high and *less* morally wrong when the frequency or magnitude of the violation was low. Ratings of moral wrongness for purity violations were *similar* regardless of whether the frequency or magnitude of the violation was low or high. Together these contrasts provide further support for the differing psychological mechanisms perspective.

Discussion

I examined whether adding weirdness to harm-based moral violations has an influence on perceived moral wrongness across variations of dosage. My results indicate that people's responses to harmful but weird violations are comparable to traditional harm violations when dosage level is varied. Moreover, harmful but weird and purity violations still produce different response patterns when dosage level is varied despite their shared weirdness. In other words, weirdness is *unable* to account for why purity, and not harm, violations are insensitive to dosage

variability. Scholars have speculated that purity-based moral violations are relatively insensitive to variations in dosage due to their inherent weirdness (Schein & Gray, 2017), a claim that my findings challenge. My data align with the mere-trace hypothesis rather than the weirdness overlap hypothesis (Rottman & Young, 2019; Wagemans et al., 2019).

My results support the differing psychological mechanisms perspective in three ways. The first is by addressing the suspected confound of weirdness. When comparing the effect size of dosage across the three violation types, harm-only and weird harm violations shared similar effect sizes (*d*s = 0.39 vs. 0.27, respectively), whereas purity violations possessed a much smaller effect size not statistically different from zero ($d = 0.02$) despite very high statistical power to detect small effects. When harm violations were modified to be weirder, they behaved in a manner to be expected had no weirdness been present at all. Had weirdness in purity violations been responsible for the insensitivity to dosage, one would expect that harmful but weird violations would have responded with a similar insensitivity to dosage as well. The second point of support results from pilot testing materials from previous studies using this paradigm. By pilot testing these materials, I am relatively confident that each moral violation was representative of the corresponding moral domain. The third point of support results from added uniformity of the dosage variations across the violation types. By ensuring that all three violation types had comparable dosage variations (e.g., "small/large" across the first item for all three violation types) I was able to avoid a potential confound of dosage inconsistency.

Limitations and Future Directions

One potential line of criticism may suggest the weirdness of purity violations compared to harm violations possess unique characteristics that evoke discomfort. In my study, harm violations were altered to be weird ("the extent to which an act is weird, bizarre, or unusual;" see

Gray & Keeney, 2015); however, these harmful but weird violations may not be *weird* in the same manner as purity violations. For example, some of the harmful but weird items could be interpreted as silly (e.g., "A person uses a Donald Duck themed magnifying glass to focus sunlight on a person's arm, burning them for 30 seconds"). Purity-based items, on the other hand, may be interpreted as more unsettling (e.g., cannibalism, incest, bestiality), serving as an indicator of one's capacity for immoral behavior. It is possible that harmful but weird acts may not fully capture what is psychologically important about purity violations. That said, several of my harmful but weird items could easily be characterized as unsettling (e.g., "A person kills five people by force feeding them cotton candy"), though examinations of individual weird-harm violations that seem silly versus unsettlingly did not reveal patterns that would support this idea.

If weirdness is not a confound that accounts for the insensitivity to dosage exhibited by purity violations, then moral psychologists should consider what aspects of purity underly this phenomenon. For instance, both harm and purity violations represent norm violations, yet many harm violations are more easily justifiable (e.g., punching someone may be self-defense) compared to purity violations (e.g., inappropriately touching a corpse; Haidt, 2012). Additionally, moral disgust acts as a "behavioral immune system" that aids in distancing oneself from situations involving disturbing sexual acts, gore, or mistreatment of carcasses (Schaller & Park, 2011). Perhaps purity-based violations provide individuals with a clear and present insight into what people are capable of thus activating this behavioral immune system. Future research may benefit from modifying harm violations to include other aspects of purity violations.

Concluding Remarks

Moral judgments are influenced by considerations of frequency and magnitude in some cases and not others. Moral judgments of situations where explicit harm has occurred rely on

indicators of frequency and magnitude more than situations where purity violations occur. My data provide evidence that weirdness is not an explanation for this pattern. These findings support a pluralistic understanding of morality suggesting the use of distinct psychological mechanisms in moral judgment dependent upon the type of moral transgression that has occurred.

References

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*(1), 1–16. https://doi.org/10.1006/obhd.1997.2690

Barr, D. J., Lévy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blanchar, J. C., Boots, J., Chambers, S., Corzine, G., Crosby, J., Daly, N., Durning, E., Lathrop, J., Looney, P., Nichols, A., Porterfield, Z., Vespole, W., Wilson, E., Xu, Y., Yabloko, V., & Zheng, T. (2024). *Does dosage distinguish impurity from harm in moral judgments? A replication and extension of Rottman and Young (2019)*. [Unpublished manuscript], Psychology Department, Swarthmore College.

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, *47*(4), 1178–1198. https://doi.org/10.3758/s13428-014-0551-2

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

Fitouchi, L., André, J., & Baumard, N. (2022). Moral disciplining: The cognitive and evolutionary foundations of puritanical morality. *Behavioral and Brain Sciences*, *46*(293), 1–74. https://doi.org/10.1017/s0140525x22002047

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. https://doi.org/10.1037/a0015141

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The pragmatic validity of moral pluralism. In *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (2018). Moral Foundations Theory: On the advantages of moral pluralism over moral monism. In K. Gray & J. Graham (Eds.), Atlas of Moral Psychology (pp. 211–222). New York: Guilford.

Gray, K., & Keeney, J. E. (2015). Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychological and Personality Science*, *6*(8), 859–868. https://doi.org/10.1177/1948550615592241

Gray, K., DiMaggio, N., Schein, C., & Kachanoff, F. (2023). The problem of purity in moral psychology. *Personality and Social Psychology Review, 27*(3), 272–308. https://doi.org/10.1177/10888683221124741

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517–523. https://doi.org/10.1016/S1364-6613(02)02011-9

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. https://doi.org/10.1037/0022-3514.65.4.613

Haidt, J., Rozin, P., Mccauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology and Developing Societies*, *9*(1), 107–131. https://doi.org/10.1177/097133369700900105

Haidt, J., Björklund, F., & Murphy, F. (2000). Moral dumbfounding: When intuition finds no reason. [Unpublished Manuscript]. Psychology Department, University of Virginia. https://polpsy.ca/wp-content/uploads/2019/05/haidt.bjorklund.pdf

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

Haidt, J. (2003). The moral emotions. In: R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford University Press. https://psycnet.apa.org/record/2009-07773-044

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

Haidt, J. (2012). The righteous mind: Why good people are divided by politics and religion. *Pantheon Books.*

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

Hume, D. (1777). An enquiry concerning the principles of morals. In *Oxford University Press eBooks* (pp. 167–169). https://doi.org/10.1093/oseo/instance.00046351

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In *Elsevier eBooks* (pp. 151–235). https://doi.org/10.1016/b978-0-12-498640-4.50011-1

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LMERTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lenth, R. (2023). EMMEANS: Estimated marginal means, aka least-squares means. [Dataset]. https://doi.org/10.32614/cran.package.emmeans

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An r package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139. https://doi.org/10.21105/joss.03139

Olatunji, B. O., Haidt, J., McKay, D., & David, B. (2008). Core, animal reminder, and contamination disgust: Three kinds of disgust with distinct personality, behavioral, physiological, and clinical correlates. *Journal of Research in Personality*, *42*(5), 1243–1259. https://doi.org/10.1016/j.jrp.2008.03.009

Piaget, J. (1932). *The moral judgment of the child.*

R Core Team (2023). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/

Rottman, J., & Young, L. (2019). Specks of dirt and tons of pain: Dosage distinguishes impurity from harm. *Psychological Science*, *30*(8), 1151–1160. https://doi.org/10.1177/0956797619855382

Rozin, P., Haidt, J., & McCauley, C. R. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 757–776). Guilford Press.

Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, *20*(2), 99–103. https://doi.org/10.1177/0963721411402596

Schein, C., & Gray, K. (2017). The Theory of Dyadic Morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–70. https://doi.org/10.1177/1088868317698288

Smith, M. A. (2018). Three kinds of moral rationalism. In *Oxford University Press eBooks*. https://doi.org/10.1093/oso/9780198797074.003.0003

Sweetman, J., & Newman, G. A. (2020). Replicating different roles of intent across moral domains. *Royal Society Open Science*, *7*(5), 190808. https://doi.org/10.1098/rsos.190808

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*(7), 320–324. https://doi.org/10.1016/s1364-6613(03)00135-9

Uhlmann, E. L., & Zhu, L. (2013). Acts, persons, and intuitions. *Social Psychological and Personality Science*, *5*(3), 279–285. https://doi.org/10.1177/1948550613497238

Wagemans, F. M. A., Brandt, M. J., & Zeelenberg, M. (2018). Disgust sensitivity is primarily associated with purity-based moral judgments. *Emotion*, *18*(2), 277–289. https://doi.org/10.1037/emo0000359

Wagemans, F. M. A., Brandt, M. J., & Zeelenberg, M. (2019). Weirdness of disgust sensitivity items predicts their relationship to purity moral judgments. *Personality and Individual Differences*, *146*, 182–187. https://doi.org/10.1016/j.paid.2018.07.042

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*(2), 202–214. https://doi.org/10.1016/j.cognition.2011.04.005

Table 1
Pilot Study: Linear Mixed Effects Models

Outcome	Model	*F*	*df*	*p*
Harm Ratings	Type	11.35	2, 27.33	<.001
	Dosage	44.75	1, 2195.13	<.001
	Type × Dosage	6.76	2, 2199.73	.001
Purity Ratings	Type	11.26	2, 27.28	<.001
	Dosage	41.15	1, 2189.63	<.001
	Type × Dosage	2.00	2, 2193.65	.136
Weirdness Ratings	Type	18.54	2, 27.05	<.001
	Dosage	47.86	1, 2192.60	<.001
	Type × Dosage	10.10	2, 2197.47	<.001

Note. N = 168.

Table 2
Pilot Study: Planned Contrasts

Outcome	Dosage	Contrast	*b* (*SE*)	95% CI	*t*	*p*
Harm Ratings	Low	Harm - Purity	0.96 (.29)	0.22, 1.70	3.28	.008
		Harm - WeirdHarm	0.27 (.30)	-1.02, 0.48	-0.92	>.999
		Purity - WeirdHarm	1.23 (.29)	-1.98, -0.49	- 4.19	<.001
	High	Harm - Purity	1.37 (.29)	0.63, 2.11	4.69	<.001
		Harm - WeirdHarm	0.26 (.29)	-0.48, 1.01	0.90	>.999
		Purity - WeirdHarm	1.10 (.29)	-1.84, -0.36	-3.78	.002
Purity Ratings	Low	Harm - Purity	-1.15 (.27)	-1.84, -0.47	-4.25	<.001
		Harm - WeirdHarm	-0.09 (.27)	-0.79, 0.60	-0.35	>.999
		Purity - WeirdHarm	1.06 (.27)	0.37, 1.75	3.88	.002
	High	Harm - Purity	-0.93 (.27)	-1.61, -0.25	-3.45	.005
		Harm - WeirdHarm	0.20 (.27)	-0.49, 0.88	0.72	>.999
		Purity - WeirdHarm	1.12 (.27)	0.44, 1.81	4.17	<.001
Weirdness Ratings	Low	Harm - Purity	-1.88 (.29)	-2.61, -1.15	-6.51	<.001
		Harm - WeirdHarm	-1.56 (.29)	-2.30, -0.83	-5.39	<.001
		Purity - WeirdHarm	0.31 (.29)	-0.42, 1.04	1.09	.868
	High	Harm - Purity	-1.32 (.29)	-2.05, -0.60	-4.60	<.001
		Harm - WeirdHarm	-1.06 (.29)	-1.79, -0.33	-3.68	.003
		Purity - WeirdHarm	0.26 (.29)	-0.47, 0.99	0.90	>.999

Note. N = 168. Bonferroni method used for p value adjustments.

Table 3
Study: Descriptive Statistics

Domain	**Dosage**	***M***	***SD***	**95% CI**
Harm	High	80.19	24.38	74.1, 85.6
	Low	68.45	29.65	63.0, 74.5
Purity	High	75.27	31.87	68.9, 80.4
	Low	73.37	31.61	68.2, 79.8
Weird-Harm	High	81.88	24.19	75.5, 87.0
	Low	72.73	29.32	67.6, 79.1

Note. N = 314. All items were rated on a 100-point scale from 0 (not at all morally wrong) to 100 (extremely morally wrong) response scales.

Figure 1

Hypothetical results supporting the differing psychological mechanisms perspective

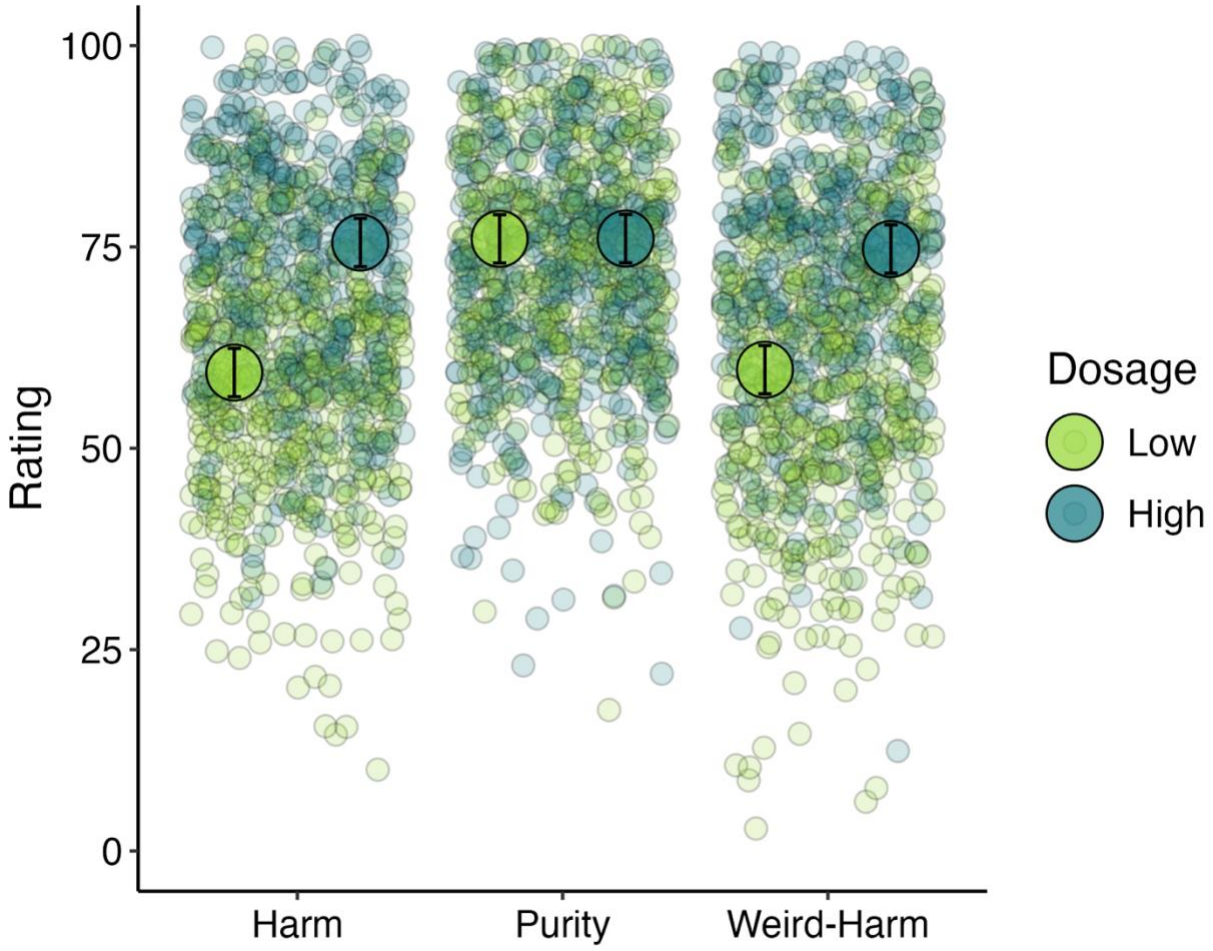


Figure 2

Hypothetical results supporting the weirdness as a confound, harm dependent perspective

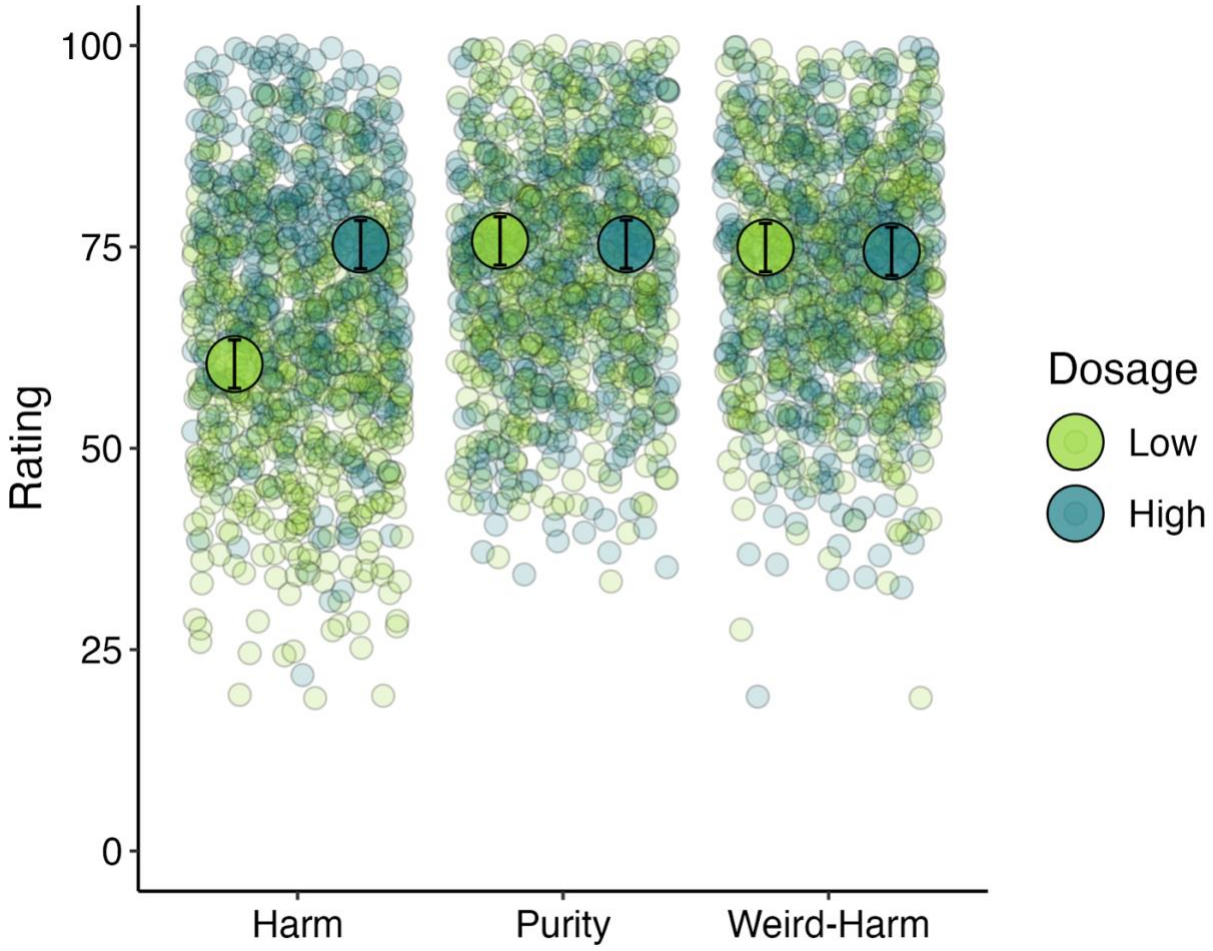


Figure 3
A priori power analysis varying sample size via the 'simr' R package

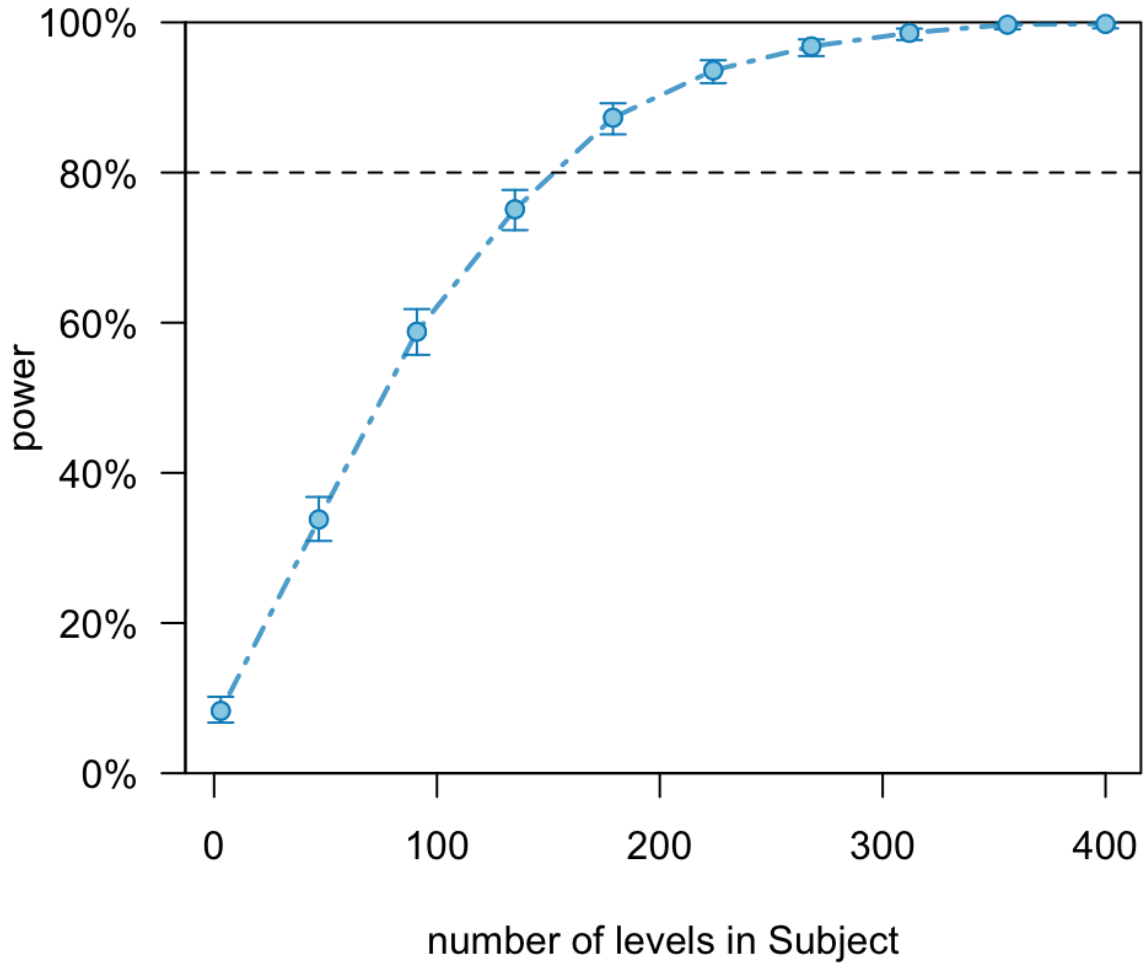
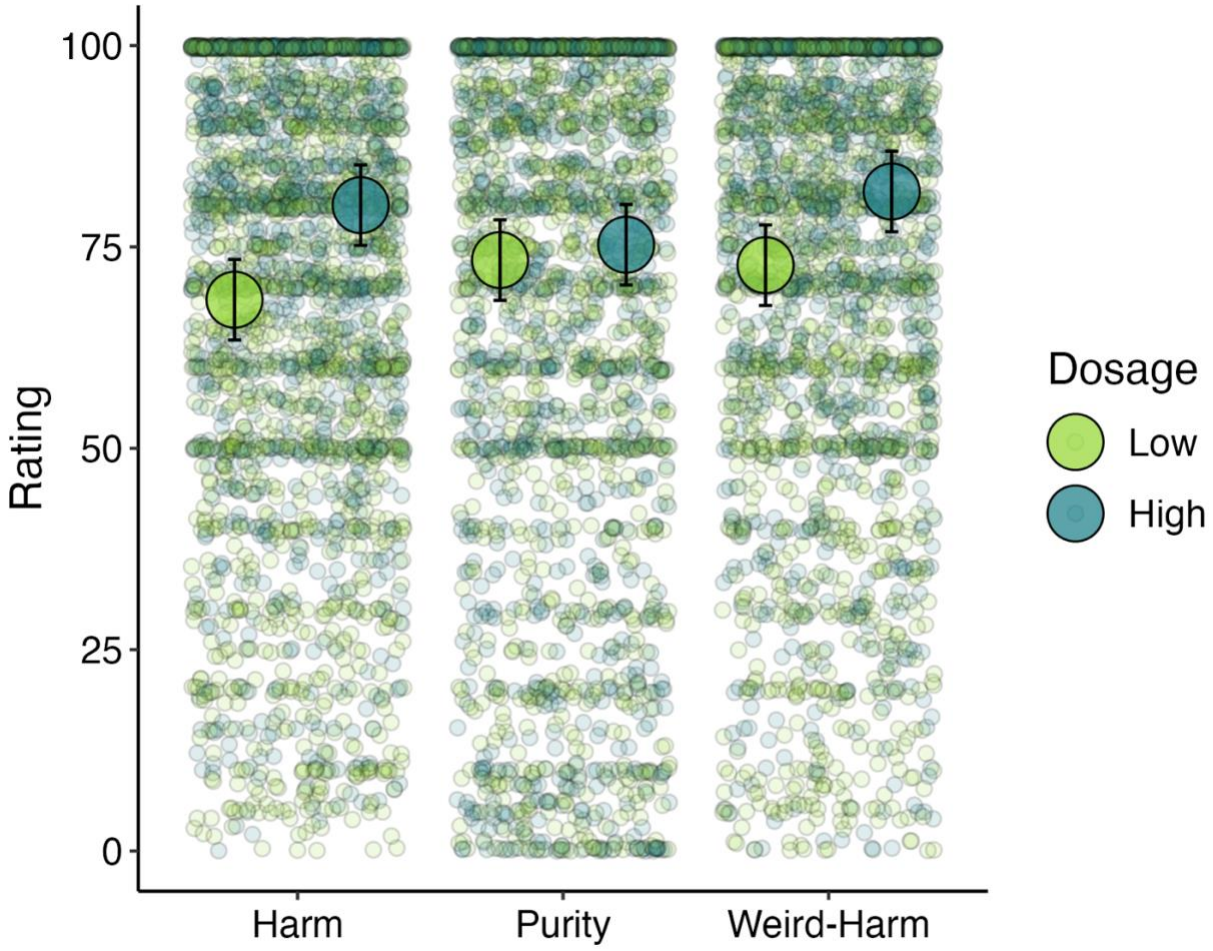


Figure 4
Results from Study



APPENDICES

Appendix A
Documentation of Preregistration

Purity, Harm, and Weird-Harm Morality (#201403)

Author(s)
This pre-registration is currently anonymous to enable blind peer-review.
It has 2 authors.

Pre-registered on: 11/26/2024 12:50 PM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Are moral judgments of purity, harm, and weird-harm violations equally sensitive to variations of dosage (i.e., frequency and magnitude)?

3) Describe the key dependent variable(s) specifying how they will be measured.
Participants will indicate the moral wrongness of 10 harm violations, 10 purity violations, and 10 harmful but weird violations in random order. Following Rottman and Young (2019), moral judgments will be measured with the following question: "How morally wrong was this action?" 0 (not at all) to 100 (extremely). Also following Rottman and Young (2019), participants will respond to two attention check items: "A person destroys the entire planet" and "A person gives money to a charitable organization" (1 = not at all immoral, 100 = extremely immoral)

4) How many and which conditions will participants be assigned to?
This study design is a 3 (Moral Domain/Type: Harm, Purity, Weird-Harm) X2 (Dosage: Low, High) factorial experiment all within-subjects.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Moral judgments will be analyzed by fitting a linear mixed-effects model with moral domain/type (harm, purity, weird-harm), dosage (low, high), and their interaction as fixed-effects predictors. I will begin by fitting the model with the maximal random-effects structure that includes random intercepts and slopes of participants and items. The random-effects structure will be reduced in complexity as needed to allow the model to converge and to avoid singular fit. Based on Rottman and Young (2019), a random-intercepts only structure is expected. Planned simple effects tests with Bonferroni correction will follow to probe interactions by comparing low vs. high dosage conditions for each moral violation type (harm, purity, and weird-harm). Furthermore, planned contrasts will test whether differences in dosage (low vs. high) differ between domain conditions (i.e., harm vs. purity, harm vs. weird-harm, and purity vs. weird-harm)

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Individuals that report being under 18 years of age, a non-native English speaker, those not living within the United States of America at the time of participation will be excluded. Additionally, participants that failed either of the two attention checks, those responding with less than 50 to "A person destroys the entire planet" or higher than 50 for "A person gives money to a charitable organization," will be excluded from the analysis as well.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
A sample of 330 participants will be recruited via Prolific's online platform.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Not applicable.

Appendix B
IRB Determination

Twin Cities Campus

Human Research Protection Program
Research and Innovation Office

350-2 McNamara Alumni Center
200 Oak St. SE
Minneapolis, MN 55455

612-626-5654
irb@umn.edu
https://research.umn.edu/units/irb

EXEMPTION DETERMINATION

Today's Date:	September 24, 2024
Investigator:	John Blanchar
IRB ID:	STUDY00023285
Study Title:	Judgments of Moral Appropriateness
Determination Date:	9/24/2024

The IRB reviewed the above submission and determined that this study meets the criteria for exemption from IRB review.

This study met the following category for exemption:

• (2) Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) and the following criteria are met: (i) The information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects (ii) Any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, educational advancement, or reputation

Ongoing Investigator Responsibilities

Ongoing IRB review and approval for this study is not required; however, this determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a Modification to the IRB for a determination.

- **Study Conduct:** You are required to follow the requirements listed in the Investigator Manual (HRP-103), available here: Toolkit Library
- **Changes in study personnel:** All study personnel must be listed on your ETHOS application and are required to completed human research training
- **Significant changes to study design:** Significant changes to your study design must be submitted to determine if your research continues to fall within the regulatory definition of "Exempt"
- **Study Closure:** You must notify the IRB when your research is complete so the record can be closed

You may have other University obligations and requirements. You are responsible for obtaining any additional approvals required for this project.

Exempt determinations are active for a period of two years. At the end of the project, but no longer than two years from the initial IRB exempt determination date, you must close this exempt study in ETHOS. If you wish to continue your research beyond two years, you will be required to submit a new exemption request for IRB consideration.

To arrive at this determination, the IRB used the worksheet HRP-312 - Exemption Determination. If you have any questions about this determination, please review this worksheet in the HRPP Toolkit Library and contact the IRB office if needed.

Study Information and Documentation			
Often sponsors or agencies request formal documentation regarding the IRB's FWA and registration. Use this letter for these requests.			
Funding:	None	Proposal/ Award ID:	None
Internal UMN Funding:	Departmental funding : The Psychology Department at the University of Minnesota Duluth is funding this research.	Fund Management Outside University:	None
Documents Reviewed with this Submission:	• AdvStudy 1, Pilot, 2A, 3A 3B.pdf, Category: Recruitment Materials; • Consent for Pilot Study of Study 1, Category: Consent Form; • Consent Form for Studies 2A and 3A, Category: Consent Form; • Consent Form for Study 1, Category: Consent Form; • hrp-580, Category: IRB Protocol; • Information Consent Sheet - Study 2B.pdf, Category: Consent Form; • Moral Appropriatness Materials - Study 1, pilot, 2A, 2B, 3A, & 3B.pdf, Category: Other; • Verbal Consent Script for Study 2B (updated), Category: Consent Form;		

Sincerely,

Angela Gerend
IRB Analyst, MA

We strive to provide clear, consistent and timely service to maintain a culture of respect, beneficence and justice in research. Complete a brief survey about your experience.

Appendix C
Prolific Recruitment Advertisements

Preliminary stimulus pilot testing data collection for use in Study

Title: Judgments of Moral Appropriateness

Description: This study examines moral judgment. To participate in this study, you must be 18 years of age or older, native English speaking, and live in the United States of America. You will be asked to complete a short online questionnaire, including questions about your background and your judgments about various actions described in short scenarios. We expect the questionnaire to take 10 minutes to complete, and participants will receive $2 as compensation.

Study

Title: Judgments of Moral Appropriateness

Description: This study examines moral judgment. To participate in this study, you must be 18 years of age or older, native English speaking, and live in the United States of America. You will be asked to complete a short online questionnaire, including questions about your background and your judgments about the moral appropriateness of various actions described in short scenarios. We expect the questionnaire to take 6 minutes to complete, and participants will receive $1.20 as compensation.

Appendix D
Preliminary Stimulus Pilot Testing Consent Form

Judgments of Moral Appropriateness

You are invited to participate in a research pilot study of moral judgment of different hypothetical moral violations. You were selected as a possible participant because you are 18 years of age or older, native English speaking, and live in the United States of America. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by Jared Boots, Department of Psychology, University of Minnesota Duluth.

Activities:

If you agree to be in this study, we would ask you to do the following things:

You will be asked to complete an online questionnaire, including questions about your background and perceived weirdness, harmfulness, and impurity of presented scenarios. We expect that you will be in this research study for 10 minutes today.

Confidentiality:

During the project, information from this study will be kept private and will be stored securely. Only the research team will have access to information that identifies you. Your identifying information will not be shared with others outside of this research study. However, organizations that may inspect and copy your information include the Institutional Review Board (IRB), the committee that provides ethical and regulatory oversight of research, and other representatives of this institution, including those that have responsibilities for monitoring or ensuring compliance (such as the Quality Assurance Program of the Human Research Protection Program (HRPP)).

Any personal information that could identify you will be removed or changed before we publish any report or share the results or data from this study. Data collected in this study may be made available in a public data repository for others to use, including for future research studies on similar or different topics, teaching, or other purposes.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to withdraw at any time without affecting those relationships.

Will I be compensated for my participation?

If you agree to take part in this research study, we will pay you $2 for the completion of the online questionnaire for your time and effort.

To Contact the Research Team:

The researcher(s) conducting this study is (are): Jared Boots. You may ask any questions you have prior to participating by contacting the lead investigator. **If you have questions later regarding the study or your participation**, **you are encouraged to contact the research team at the University of Minnesota Duluth, 320-287-1273, boots027@morris.umn.edu.**

To Contact Someone Outside of the Research Team:

This research has been reviewed and determined exempt by the Institutional Review Board (IRB), which is part of the Human Research Protections Program (HRPP). If there is an issue you would like to discuss with someone who is not on the research team you are encouraged to call the HRPP Research Participants' Advocate Line at 612-625-1650 (Toll Free: 1-888-224-8636) or go to z.umn.edu/participants. For example:

- Your questions, concerns, or complaints are not being answered by the research team.
- You are having difficulty reaching the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research participant.
- You want to provide feedback about this research to someone who is not on the study team.

You will be given a copy of this information to keep for your records.

Appendix E
Preliminary Stimulus Pilot Testing Demographics Questionnaire

What is your prolific ID?
Please note that this response should auto-fill with the correct ID
[Type response]

What is your age?
[Type response]

Race (choose as many as apply)
- Black or African American: A person having origins in any of the black racial groups of Africa.
- American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.
- Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent; for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific islands.
- White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.

What is your gender?
- Male
- Female
- Neither (please specify)
  [Type response]

Is English your primary/first language?
- Yes
- No

Appendix F
Preliminary Stimulus Pilot Testing Materials

Thank you for agreeing to take participate in today's study. Please read the following scenarios and indicate your responses. There are no right or wrong answers.

Preliminary stimulus pilot testing questions- "How *weird (i.e. unusual, bizarre, odd)* is this action?", "How *harmful (i.e., physically or emotionally damaging)* is this action?", "How *impure (i.e., dirty, tainted, or taboo)* is this action?" using 1 (*not at all weird/harmful/impure*) to 7 (*extremely weird/harmful/impure*) response scales.

	Harmful but Weird
1	A person feeds a mongoose broken glass [once/frequently].
2	A person intimidates a coworker by writing cryptic, threatening messages in Morse code on their computer screen [every so often/every week].
3	A person intentionally gives a sibling a concussion with a frozen fish on [one occasion/many occasions].
4	A person throws a [tiny/large] microwave at a chicken.
5	A person slaps [another person/20 other people] with a handful of sparkly glitter.
6	A person aggressively gives somebody a [tiny/huge] welt on their arm using a holiday themed nutcracker with googly eyes.
7	A person jabs a candy cane [one millimeter/one centimeter] into the skin of her brother's leg.
8	A person jabs a classmate with the tail of a purple toy dinosaur and draws [a single drop/a teaspoon] of blood.
9	A pharmacist gives her patients a drug that contains [0.001%/0.1%] of rocket fuel.
10	A person strangles his sibling with several candy licorice ropes for [a split second/30 seconds].
11	A person secretly fills a café coffee pot with [one gram/a quarter pound] of two-in-one shampoo and conditioner.
12	A person removes a live shrimp from its tank, denying it of water, and places it on a vibrating phone for [an instant/a minute].
13	A person captures [a single worm/10 worms] and makes [it/them] sick with lead-contaminated pickle juice.
14	A father smacks his child with a rubber chicken [once/repeatedly].
15	A person kills [one person/five people] by force feeding them cotton candy.
16	A person coats a cat in glue [once/frequently].
17	A person insults an obese person by comparing his or her resemblance to a mythical creature [once/frequently].
18	A person stomps on a burrowed rabbit nest with honking clown shoes, killing [one/10] rabbit[s].
19	A man starts [one/10] malicious rumor[s] that another person enjoys giving stray cats tattoos.
20	A girl lies to a boy on [one occasion/ several occasions] that his girlfriend cheated her way through college by bribing professors with coupons to the circus.

21	A boy starts a rumor on [one occasion/ several occasions] that his classmate believes himself to be an octopus.
22	A boy scares his sister on [one occasion/ several occasions] by telling her that if she watches too much television, she will turn into a subterranean mole person.
23	A man starts [one/10] rumor[s] that another person has a history of abusing stuffed animals with food condiments.
24	A man starts a rumor that another man is secretly a lizard person [one occasion/several occasions].
25	A boy lies to a girl on [one occasion/ several occasions] that her boyfriend runs a church for clowns.
26	A woman lies to her boss on [one occasion/several occasions] that one of her coworkers swaps her lunch with plastic toy food items.
27	A girl starts a rumor on [one occasion/ several occasions] that another student enjoys eating raw eggs with a straw.
28	A girl lies to her mother [once/several times] that her sister stole the neighbor's toilet seat.
29	A woman starts a rumor that another worker swaps out candy from the breakroom with fake candy on [one occasion/several occasions].
30	A girl tosses [one/10] commemorative birthday bowling pin(s) at chickens in their coop.
31	A teenager releases [one/10] giant fiddler crabs to kill raccoons in his neighborhood.
32	A coach hits a player's hand [once/five times] with a can of expired refried beans for not playing hard.
33	A man slap another man with a warm Belgian waffle [once/five times].
34	A boy dumps a [small/large] hot cup of cough syrup on another child.
35	A girl smacks her boyfriend on the back [once/repeatedly] with a frozen burrito.
36	A person robs [one/six] gas station[s] wielding a live alligator.
37	A person trains their dog to bite (the first person/ the first five people) it sees wearing yellow clothing.
38	A person puts a turtle with a taser mounted on its back in their neighbor's mailbox on [one occasion/many occasions].
39	A person puts [one/ several] shark tooth(teeth) covered in baby oil in a person's shoes.
40	A person uses a Donald Duck themed magnifying glass to focus sunlight on a person's arm, burning them for [three seconds/10 seconds].

	Harm	**Purity**
1	A person starves a goat [once/frequently].	A person has intercourse with a goat [once/frequently].
2	A person bullies a colleague [every so often/every week].	A person uses heroin [every so often/every week].
3	A person [occasionally/regularly] holds meetings in a room with high asbestos levels.	A person [occasionally/regularly] masturbates with a chicken carcass.
4	A person purposely knocks a sibling unconscious on [one occasion/many occasions].	A person has protected sex with a sibling on [one occasion/many occasions].
5	A person throws a [small/large] rock at a farm animal.	A person eats a [small/large] amount of flesh from a dead person.
6	A person punches [one person/20 people].	A person inappropriately touches [a corpse/20 corpses].
7	A person angrily gives somebody a [tiny/huge] bruise on their leg.	A person builds a [tiny/huge] house on ground that is considered sacred.
8	A person pushes a thumbtack [one millimeter/one centimeter] into the skin of her classmate's back.	A woman consumes a [one-millimeter-/one-centimeter-] thick piece of flesh from her neighbor's corpse.
9	A person pricks a colleague with a knife and draws [a single drop/a teaspoon] of blood.	A person pours [a single drop/a teaspoon] of sterile urine into their colleague's drink.
10	A doctor prescribes her patients a medication that contains a [0.001%/0.1%] trace of mercury.	A person engineers the genome of a monkey embryo to make it [0.001%/0.1%] more similar to a human.
11	A boy puts his sister in a chokehold for [a split second/thirty seconds].	A boy lets his sister fondle his genitals for [a split second/ thirty seconds].
12	A person secretly puts [one gram/a quarter pound] of hot sauce into a restaurant's ketchup bottle.	A person puts [one gram/a quarter pound] of feces into an urn of their grandmother's ashes.
13	A boy takes his pet fish out of its bowl, depriving it of water, for [an instant/a minute].	A person whispers vulgar obscenities into a priest's ear for [an instant/a minute].
14	A woman sprays [one particle/a fluid ounce] of a hazardous chemical into a residential area.	A person sprays [one particle/a fluid ounce] of cow's blood onto a holy wooden crucifix.
15	A mother slaps her child [once/repeatedly].	A person puts [a single ladybug/10 ladybugs] onto their plate and eats [it/them].
16	A person kills [one person/5 people].	A man has sex with a frozen chicken [Once/frequently] before cooking it for dinner.
17	A person kicks a dog in the head [once/frequently].	A very drunk woman makes out with [a single/several] stranger(s) on the city bus.
18	A person makes cruel remarks to an overweight person about his or her appearance [once/frequently].	A family eats the carcass(es) of [one/three] of their pet dogs that had been run over.
19	A girl steps on an ant hill, killing [hundreds/thousands] of ants.	A drunk elderly man offers to have oral sex with [one person/several people] in the bar.
20	A person sticks [a pin/several pins] into the palm of a child they don't know.	A person urinates in the wave pool at a [crowded/non-crowded] amusement park.

Appendix G
Study Demographics Questionnaire

What is your prolific ID?
Please note that this response should auto-fill with the correct ID
[Type response]

What is your age?
[Type response]

Race (choose as many as apply)
- Black or African American: A person having origins in any of the black racial groups of Africa.
- American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.
- Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent; for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific islands.
- White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.

What is your gender?
- Male
- Female
- Neither (please specify)
  [Type response]

Is English your primary/first language?
- Yes
- No

Appendix H

Study Materials

Thank you for agreeing to take participate in today's study. Please read the following scenarios and indicate your responses. There are no right or wrong answers.

Study questions- "How morally wrong was this action?" using 0 (*not at all morally wrong)* to 100 (*extremely morally wrong)* response scales.

	Harm	**Purity**	**Harmful but Weird**
1	A person throws a [small/large] rock at a farm animal.	A person eats a [small/large] amount of flesh from a dead person.	A child dumps a [small/large] hot cup of cough syrup on another child.
2	A person punches [one person/20 people].	A person inappropriately touches [a corpse/20 corpses].	A person captures[s] [a single/20] hungry bear[s] from a bicycle riding circus act and releases [it/them] inside an elementary school.
3	A person purposely knocks a sibling unconscious on [one occasion/many occasions].	A person has protected sex with a sibling on [one occasion/many occasions].	A person gives their sibling a concussion with a frozen fish on [one occasion/many occasions].
4	A person angrily gives somebody a [tiny/large] bruise on their leg.	A person puts a [tiny/large] amount of feces into an urn of their grandmother's ashes.	A person throws a [tiny/large] microwave at a chicken.
5	A person bullies [one/five] colleague[s].	A family eats the carcass of [one/five] of their pet dogs that had been run over.	A person robs [one/five] gas stations wielding a live alligator.
6	A person makes cruel remarks to an overweight person about his or her appearance [once/ frequently].	A man has sex with a frozen chicken [once/frequently] before cooking it for dinner.	A person insults an obese person [once/frequently] by comparing their resemblance to that of a mythical creature.
7	A boy puts his sister in a chokehold for [an instant/thirty seconds].	A person whispers vulgar obscenities into a priest's ear for [an instant/30 seconds].	A person uses a Donald Duck themed magnifying glass to focus sunlight on a person's arm, burning them for [an instant/30 seconds].
8	A person kills [one person/five people].	A person pours sterile urine into [one/five] of their colleague['s/s'] drink.	A person kills [one person/five people] by force feeding them cotton candy.
9	A mother slaps her child [once/repeatedly].	A person sprays cow's blood onto a holy wooden crucifix [once/repeatedly].	A coach hits a player's hand [once/repeatedly] with a can of expired refried beans for not playing hard.
10	A person [occasionally/regularly] holds meetings in a room with high asbestos levels.	A very drunk woman [occasionally/regularly] makes out with strangers on the city bus.	A person [occasionally/regularly] puts a turtle with a strong taser mounted on its back in their neighbor's mailbox.

Appendix I
Preliminary Stimulus Pilot Testing and Study Attentions Check Items

Thank you for agreeing to take participate in today's study. Please read the following scenarios and indicate your responses. There are no right or wrong answers.

1- "A person destroys the entire planet."
2- "A person gives money to a charitable organization."

Preliminary stimulus pilot testing questions- "How *weird (i.e. unusual, bizarre, odd)* is this action?", "How *harmful (i.e., physically or emotionally damaging)* is this action?", "How *impure (i.e., dirty, tainted, or taboo)* is this action?" using 1 (*not at all weird/harmful/impure*) to 7 (*extremely weird/harmful/impure*) response scales.

Study questions- "How morally wrong was this action?" using 0 (*not at all morally wrong)* to 100 (*extremely morally wrong)* response scales.