Where (And When) Is The Harm? Moral Judgments Under Time Pressure

Matthew E. Vanaman, Hanah A. Chapman

Brooklyn College, City University of New York

Abstract

Recently, moral psychologists have debated whether concerns about harm cause condemnation of immoral behaviors, even when those behaviors are harmless. For example, behaviors that violate *purity*, or moral concerns that regulate food, sex, and hygiene, are routinely condemned despite their harmlessness. Moral Foundations Theory (MFT) proposes that while harmful behaviors are indeed condemned because they cause harm, condemnation of purity behaviors is driven by disgust. MFT also argues that any harm concerns expressed toward purity behaviors are post-hoc rationalizations offered to persuade others. By contrast, the Theory of Dyadic Morality (TDM) proposes that purity behaviors are perceived to be harmful, and that people “fill in” the absent victim who is harmed. In the current research, I tested these hypotheses using time pressure to limit participants’ capacity for effortful reasoning as they made moral judgments about harm and purity behaviors. If disgust drives condemnation of purity behaviors (per MFT), and harm concerns are post-hoc rationalizations, then people should show concerns about harm less frequently when judging impure behaviors under time pressure. If harm drives condemnation of purity behaviors (per TDM), then time pressure should have no effect on or even increase the frequency of harm concerns. Neither MFT nor TDM predicts that judgments about harm behaviors should be affected by time pressure. I found that time pressure had no effect on the frequency of harm concerns when judging purity behaviors, providing some evidence against MFT. Unexpectedly, this study found that time pressure led participants to more frequently endorse disgust as the reason they thought harmful behaviors were immoral. This finding contradicts both the MFT and TDM perspectives but fits with some previous research associating disgust with moral concerns beyond purity. This finding provides a potentially exciting avenue for the integration of disgust into research on non-purity moral values.

Where (And When) Is The Harm? Moral Judgments Under Time Pressure

People sometimes insist that certain immoral acts are harmful, even when the acts do not result in harm. For example, when asked who might be victimized by someone urinating on a tombstone, the vast majority of those who found urinating on a tombstone immoral identified at least one entity would be harmed (e.g. the family of the deceased or the deceased themselves; DeScioli, Gilbert, & Kurzban, 2012). Theorists in moral psychology have attempted to understand why this happens. Moral Foundations Theory (MFT) argues that morality evolved to curtail behaviors which are counterproductive to reproductive fitness, and that innate *moral foundations* cause social groups to moralize and regulate the kinds of behaviors that might have been evolutionarily maladaptive without regulation (Haidt & Joseph, 2007). MFT posits six such foundations: Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, Purity/Degradation, and Liberty/Oppression (Graham et al., 2013). To date, most of the evidence for MFT has come from work contrasting the foundations of Care/Harm (henceforth harm), or protection of the weak and vulnerable; and Purity/Degradation (henceforth purity), or protecting the sacred from material or spiritual corruption. MFT argues that condemnation of harm is driven by the attachment system, which evolved to alert parents to signs of suffering in their offspring and became more generalized over time. Therefore, MFT posits that harm concerns drive condemnation of harmful acts. Purity, on the other hand, is proposed to have evolved out of the disease avoidance system, resulting in the moralization of harmless but disease-threatening behaviors (Graham et al., 2013). Purity, as such, is driven by disgust, the primary disease avoidance emotion (Oaten, Stevenson, & Case, 2009; Clifford & Wendell, 2016; Haidt et al., 1993; Rottman et al., 2014; Scott, Inbar, & Rozin, 2016; Seidel & Prinz, 2013; Wagemans, Brandt, & Zeelenberg, 2018).

Although MFT posits that condemnation of purity behaviors is driven by disgust, people nonetheless frequently report that purity behaviors are harmful or have a victim, even when this is implausible. To account for this, MFT argues that stated concerns about harm actually reflect a post-hoc rationalization offered to persuade others to share one’s intuition (Haidt, 2012). Preliminary evidence for this perspective is found in Haidt, Bjorklund, and Murphy (2000), where researchers asked participants whether several purity behaviors, such as safe, consensual sexual relations between siblings, are immoral and why. They found that despite initially arguing that the acts were harmful, some participants continued to condemn the behaviors after eventually agreeing that the behaviors were victimless and safe. The researchers concluded that the initial concerns about harm thus could not have been the cause of the condemnation. A handful of other studies have supported the post-hoc rationalization view by showing that condemnation of harmless-but-impure behaviors is present even after statistically controlling for concerns about harm to the self, God, or friends and family.

This perspective is not without its criticisms. The Theory of Dyadic Morality (TDM) has offered its own evidence that concerns about harm drive people to condemn harmless purity behaviors. According to TDM, all moral condemnation is driven by recognition of a moral agent (a harm-doer) causing harm to a moral patient (a harmed victim). This moral “dyad” is necessary for behaviors to be considered immoral, even in the absence of identifiable harmful consequences (Schein & Gray, 2017). Taking a constructionist approach to moral judgment rooted in theory of mind and inspired by cognitive theories of perception (Gray, Young, & Waytz, 2012), TDM is agnostic to evolutionary explanations of behavior. Instead, it is primarily concerned with understanding the circumstances under which people perceive moral dyads (agent + patient pairs). TDM proposes that because people have a template for immorality rooted in agent-patient interactions, people spontaneously “fill in” for a harmed patient if they perceive an agent doing an immoral behavior. Thus, according to TDM, harm concerns toward purity are never post-hoc rationalizations, but rather reflect people's tendency to see all immoral acts as necessarily consisting of an agent and patient. More broadly, TDM argues against the idea of distinct moral foundations that depend on different inputs, such as harm and purity.

Evidence for TDM is seen in Schein and Gray (2015), who asked participants to rate how well foundation-specific adjectives described various behaviors of harm and purity. For example, participants rated how well the words "harmful" and "gross" described the purity behavior "eat dead dog". Across all behaviors, harm and purity adjectives were highly correlated, implying that these adjectives measure the same underlying construct of harm. Gray and Keeney (2015) used a similar method and found that harm behaviors were rated as both more harmful and more impure than purity behaviors, supporting the view that purity concerns are not distinct from harm concerns. Beyond testing distinctness, they also tested the idea that harm concerns represent post-hoc rationalizations. If harm concerns are post-hoc rationalizations, at least in the case of purity, then limiting a person's cognitive resources should in turn limit their ability to identify victims. One way to limit access to cognitive resources is to apply time pressure (Beach & Mitchell, 1978; Edland & Svenson, 1993; Kerstholt, 1994; Payne, Bettman, & Johnson, 1988; Suri & Monroe, 2003; Svenson, Edland, & Slovic, 1990; Svenson & Maule, 1993). In Gray, Schein, and Ward (2014), the authors asked participants to indicate how much they agreed that various impure behaviors caused harm. Half of the participants were randomly assigned to respond within 7 seconds; the other half of participants responded with no time pressure. Contrary to MFT's predictions, those in the time pressure condition were *more* likely to agree harmless purity behaviors caused harm, compared to those in the untimed condition. Based on these studies, proponents of TDM argue that a harm-based template is a better explanation of why people identify victims in harmless-but-impure behaviors, compared to MFT's proposal of a special disgust-based system.

While these studies appear to falsify MFT's post-hoc rationalization view, there are also flaws in their research methodology. Graham (2015) noted that relying on Likert-scale ratings of adjectives is insufficient to distinguish between moral intuitions and feelings of moral condemnation in a more general sense. For example, in Gray and Keeney (2015), harm and impurity ratings were highly correlated with severity, a measure of the strength of moral condemnation (.93 $\leq$ *r*s $\leq$ .97). This high correlation strongly suggests that ratings across all of these scales are tapping general moral condemnation. On this view, Gray and colleagues have only shown that people consider harm behaviors more immoral than purity behaviors, not that harm and purity judgments rely on the same moral judgment process. Haidt (2015) made the same critique of Schein and Gray (2015)'s adjective study: when participants rate how well the words "harm" and "gross" describe "eat dead dog", what they really want to tell you is how "immoral" they think eating dead dog is. As such, high correlations among Likert scale ratings may reveal little about the distinctness of moral concerns.

A similar critique applies to Gray et al. (2014), which showed that time pressure causes greater perceived harm. In this study, immorality and harm ratings were administered on Likert scales. Gray and colleagues may have simply shown that people find purity behaviors more immoral under time pressure and used the harm ratings to express that. This aligns with prior work showing that people initially react negatively toward behaviors of all of MFT's moral foundations, with some people becoming more permissive after deliberation (Graham et al., 2012). Perhaps most importantly, research on TDM has not shown that harm is the reason why people think purity behaviors are immoral, but simply that peoples' concerns about harm are correlated with condemnation of purity behaviors (Gray et al., 2014, p. 1608). MFT and TDM do not disagree that condemnation of purity is correlated with stated harm concerns; rather, they disagree about the temporal order of harm concerns relative to the condemnation. As such, previous research has been unable to adjudicate between these perspectives as they only assessed correlations between harm concerns and purity concerns.

In light of the methodological limitations of previous work, the current study sought to conduct a stronger test of the role of harm in condemnation of purity behaviors. To do so, this study replicated and extended Study 1 of Gray et al. (2014), which assessed the role of time pressure in participants' reporting of harm concerns. I addressed the methodological limitations of previous research in several ways. First, instead of simply asking participants how much they agree that the behaviors cause harm, as in Gray et al. (2014), this study asked participants *why* they thought the act was immoral. This prompt allows for a more face-valid interpretation of the results as representing a causal relationship between harm ratings and moral condemnation. Second, while Gray and Keeney (2015) interpreted high correlations between harm and purity measures as indicating that harm and purity are not distinct, an alternative interpretation from a measurement perspective is that harm and purity ratings simply failed to achieve discriminant validity. To address this, I implemented a forced-choice paradigm so that harm concerns were constrained to be uncorrelated with purity concerns. That is, participants cannot endorse multiple adjectives as in Schein and Gray (2015), but instead must endorse either a harm- or disgust-based reason for why an act is immoral. This forces discriminant validity between different harm and purity concerns, allowing us to isolate their respective effects.

Using this improved paradigm, I tested competing predictions about the effect of time pressure on harm judgments. For purity behaviors, MFT predicts that time pressure should lower concerns about harm, while TDM predicts that time pressure should either increase or have no effect on harm concerns. These predictions are visualized in Figure 1.

This research also added additional conditions that allow for further tests of MFT and TDM. Like Gray et al. (2014), I included a harm condition, in which subjects responded to behaviors that represent canonical behaviors of harm values. According to MFT, time pressure should not affect how frequently participants show concern about harm for harm behaviors, because harm concerns cause condemnation of harmful acts (see Figure 1). In other words, harm concerns about harmful behaviors are not post-hoc rationalizations. Thus, the harm condition provides a baseline against which to compare the purity condition, which MFT predicts should not be affected by time pressure. TDM also predicts that harm concerns about harmful behaviors should not be affected by time pressure because, like MFT, TDM proposes that harmful behaviors are condemned because they are harmful. Thus, the harm condition does not adjudicate between MFT and TDM by itself; rather, it offers a useful comparison with the purity condition, where TDM and MFT do offer competing hypotheses.

To provide an additional test of TDM, this study also included a disgusting non-moral condition consisting of norm violations that are disgusting and weird, but not immoral (e.g. "a man keeps flakes of his skin in a small container"). According to TDM, an act will be perceived as harmful only if it is negatively-valenced, high-arousal, weird, and considered immoral (Gray, Schein, & Cameron, 2017, figure 2). Specifically, participants should not perceive harm in acts that are negative, high-arousal, and weird, unless those acts are also perceived to be immoral (Schein & Gray, 2017). Gray et al. (2014) tested this hypothesis by showing that people do not see harm in sad non-moral acts (e.g. "a girl loses her favorite teddy bear"). However, this is a weak test of this hypothesis, as sadness is a low-arousal emotion (Gilet & Jallais, 2011), and the sad scenarios were non-weird. The disgusting non-moral condition addresses this concern because disgust is a high-arousal negative emotion, and I designed disgusting non-moral scenarios to be weird. TDM predicts that time pressure should not cause concerns about harm toward disgusting non-moral behaviors (Figure 1).[1] MFT is agnostic toward this condition. Thus, the disgusting non-moral condition does not distinguish against MFT and TDM; but if time pressure *does* increase harm concerns toward disgusting non-moral behaviors, that would provide evidence against TDM.

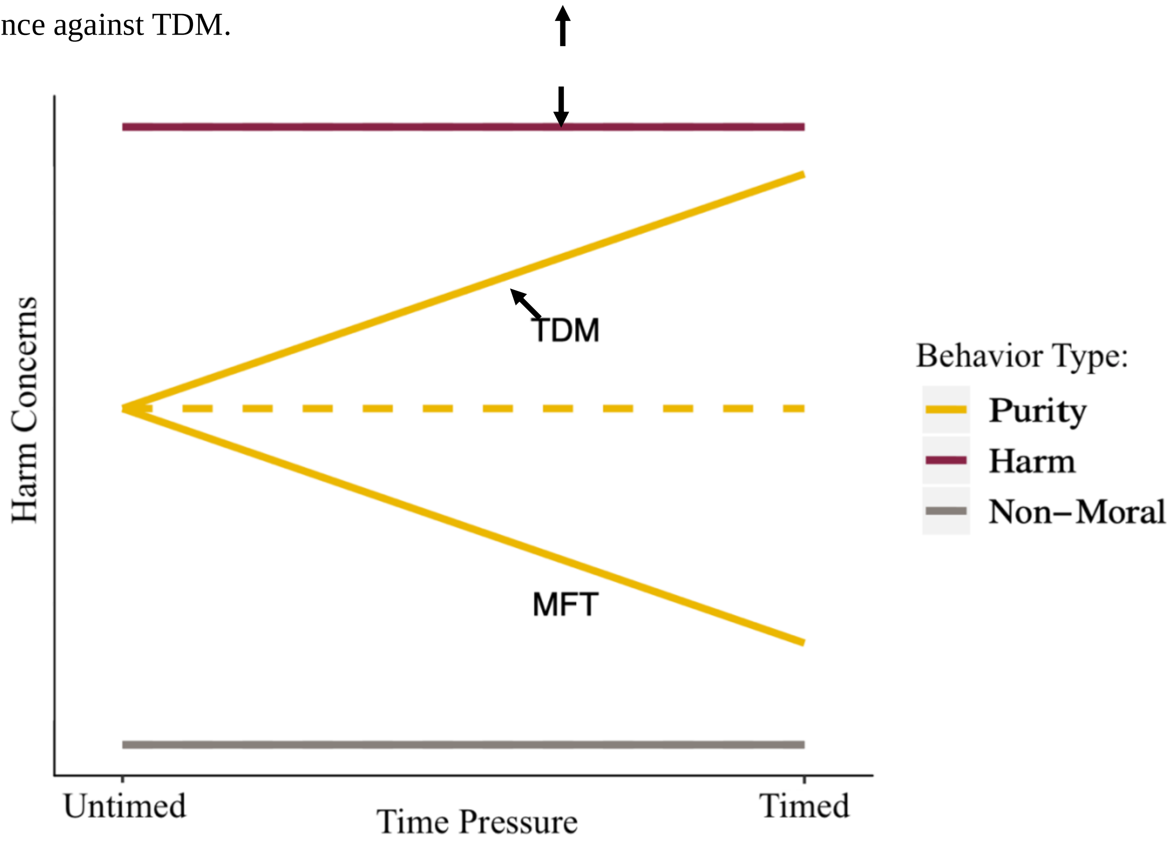


Figure 1. A visualization of the predictions of Moral Foundations Theory (MFT) and the Theory of Dyadic Morality (TDM).

Pilot Study

The first goal was to identify purity behaviors that exhibit responses that allow for statistical analysis. Because MFT predicts that time pressure should decrease the frequency of

[1] Inclusion of this condition also answers recent calls for research to compare disgusting, weird, immoral purity acts with disgusting, weird, non-moral acts (Gray & Graham, 2018, Chapter 37).

harm concerns about purity behaviors, floor effects for harm ratings in the purity condition could undermine the ability to detect such a decrease. The inverse is true for TDM. TDM predicts that time pressure should elicit the same or increased frequency of harm concerns about purity behaviors, as compared to the no-pressure condition. Ceiling effects for harm ratings in the purity condition would undermine the ability to detect such an increase. As such, the study design required purity behaviors for which participants responded with the harm concern about 50% of the time. I also sought to identify harm behaviors that were generally considered harmful and disgusting non-moral behaviors that were generally considered disgusting and weird. Finally, I sought harm and purity behaviors that were comparable in perceived immorality and disgusting non-moral behaviors that were generally low in immorality.

The second goal of piloting was to identify baseline response times for the behaviors. Doing so allowed us to select behaviors so as to minimize variability in response times as some behaviors might take longer to read than others. To accomplish this, the experiment recorded the average length of time it took a participant to choose why each behavior was immoral in the absence of time pressure.

Method

Participants

Participants in the pilot study were 28 undergraduate students recruited from the Brooklyn College institutional subject pool (M = 20.64, SD = 3.40). The sample was mostly female (46%, 10.4% transgender or non-binary), non-Hispanic (57.1%), and 17.9% white, 14.3% Asian, 14.3% black or African American, and 25% multiple race or other racial identity. Participation was given in exchange for course credit. Three participants did not finish the study, thus were excluded from analyses.

Stimuli

A preliminary pool of 32 purity behaviors (e.g., “someone has sex with a chicken before eating it”) and 35 harm behaviors (e.g., “someone purposely steps on an ant hill, killing thousands of ants”) was gathered from a review of the literature on moral judgment (e.g., Clifford, Iyengar, Cabeza, & Sinnott-Armstrong, 2015) and included several additional researcher-generated purity behaviors. Disgusting non-moral behaviors (e.g., “a man keeps flakes of his skin in a small container”; N = 35) were all researcher-generated.

Procedure

The pilot study was administered via a Qualtrics survey, which had two sections. In section 1, participants completed a moral judgment task where they read each behavior one at a time, and selected one of two reasons why they thought the behavior was immoral: “is disgusting” (henceforth *disgust option*) or “has a victim” (henceforth *harm option*; see Figure 1). Prior to beginning this task, participants were given definitions of each option. The disgust option, which mirrors MFT’s claim that condemnation of impure behaviors is driven by disgust, was defined as “bothersome or offensive because the behavior is gross, unnatural, degrading, or sinful”. The notion of something being “bothersome or offensive” was taken from Haidt et al. (1993), while the purity adjectives “gross”, “unnatural”, “degrading”, and “sinful” were taken from the Moral Foundations Dictionary (Graham, Haidt, & Nosek, 2009). The harm option was defined as “it seems like this behavior would cause someone (or something) else to experience physical or psychological pain and suffering”. This definition reflects theorizing on TDM (e.g. Gray & Wegner, 2012; Gray et al., 2012). Because the pilot study included disgusting non-moral behaviors, which I expected would be viewed as morally acceptable, I was concerned that some participants might be confused at being asked to choose why these acts were immoral. Similarly, participants might view some of the purity or harm behaviors as being morally acceptable, based on their individual values. Therefore, participants were told that they may occasionally disagree that an act is immoral and, in such cases, they should choose the option they felt best describes the behavior.

The disgust and harm options appeared below the behavior, on the same side of the screen for every behavior. Participants indicated their choice by clicking on one option. Response time was operationalized as the seconds until first click, which measures number of seconds passed from the moment the page on the computer screen opened to the moment a participant clicked their choice of options. I chose seconds until first click because a participant could only make one choice before the survey advanced, which was also true of the main experiment.

In section 2 of the pilot study, participants viewed all of the behaviors again, this time providing ratings of immorality (“how immoral is this behavior?”), disgust (“how disgusting [gross, unnatural, or strange] is this behavior?”), and weirdness (“how weird [atypical, unusual, or strange] is this behavior?”) on 5-point Likert scales ranging from 1 (not at all immoral/disgusting/weird) to 5 (extremely immoral/disgusting/weird). Participants also provided ratings of perceived harm (“Does this behavior have a victim or victims - that is, the behavior causes someone or something else to suffer physically or psychologically?”), also on a 5-point Likert scale (1 = definitely not to = definitely yes).

Results and Discussion

Forced Choice and Likert Ratings

Preliminary behavior pool. Table 1 shows summary descriptive statistics for the preliminary behavior pool. Participants responded to purity behaviors with the disgust option (i.e., indicated that the behaviors were immoral because they are disgusting) the majority of the time. Purity behaviors were also above the scale midpoints for immorality, disgust, and weirdness ratings, but below the middle of the scale for harmfulness. For harm behaviors, participants responded with the harm option (i.e. indicated that the behaviors were immoral because they had a victim) the majority of the time. Harm behaviors were similar to purity in terms of immorality and disgust, but lower in weirdness, and substantially higher in perceived harmfulness. For the disgusting non-moral behaviors, participants responded with the disgust option the majority of the time. Disgusting non-moral behaviors were seen as lower in perceived immorality and disgust than both purity behaviors and harm behaviors, lower in weirdness than purity behaviors (but higher than harm behaviors), and lower in perceived harmfulness than purity behaviors and harm behaviors.

Table 1.

Descriptive Statistics for the Preliminary Behavior Pool

Behavior Type (*n*)	Harm Option (%)	Harmfulness *M (SD)*	Immorality *M (SD)*	Weirdness *M (SD)*	Disgust *M (SD)*
Purity (32)	21.5	2.6 (1.5)	3.5 (1.5)	3.7 (1.4)	3 (1.7)
Harm (35)	76.4	4.4 (0.8)	3.2 (1.4)	2.8 (1.4)	3.3 (1.4)
Disgusting Non-Moral (35)	18.6	1.8 (1.3)	3.1 (1.4)	2.9 (1.4)	1.6 (1.2)
Overall (102)	38.8	2.9 (1.2)	3.3 (1.4)	3.1 (1.4)	2.6 (1.4)

Characteristics of retained behaviors. To avoid floor effects in the main experiment, purity behaviors were retained when 20% or more of participants responded with the harm option. Harm behaviors were retained when 80% or more of the participants responded with the harm option. Lastly, disgusting non-moral behaviors were retained when 20% or fewer of the participants responded with the harm option. Table 2. shows the final percentages of the harm option for both individual behavior types and overall (i.e. irrespective of behavior type).

Table 2. also gives means and standard deviations of the Likert-scale ratings of harmfulness, immorality, weirdness, and disgust for each behavior type.

Table *3* provides a full breakdown of results from paired-samples t-tests and effect size differences of between-behavior type mean comparisons. Compared to harm behaviors, purity behaviors were rated as substantially less harmful, moderately less immoral, moderately weirder, and roughly equally disgusting. Compared to disgusting non-moral behaviors, purity behaviors were rated as much more harmful, much more immoral, slightly less weird, and moderately less disgusting. Appendix gives proportions, means, medians, and standard deviations for individual retained behaviors, while wordings of the retained behaviors can be found in Appendix .

Table 2.

Descriptive Statistics for the Retained Behavior Set

Behavior Type (*n*)	Harm Option (%)	Harmfulness *M* (*SD*)	Immorality *M* (*SD*)	Weirdness *M* (*SD*)	Disgust *M* (*SD*)
Purity (10)	31.5	2.3 (1.4)	2.7 (1.5)	3.4 (1.4)	3.1 (1.5)
Harm (10)	82.3	4.6 (0.7)	3.4 (1.2)	2.9 (1.4)	3.1 (1.4)
Disgusting Non-Moral (10)	10.8	1.6 (1.1)	1.7 (1.3)	3.6 (1.3)	3.6 (1.3)
Overall (30)	41.5	2.8 (1.1)	2.6 (1.3)	3.3 (1.4)	3.3 (1.4)

Table 3.

Results from Paired-Samples T-Tests Among Retained Behaviors

		95% CI			
Comparison	M_{diff}	*LB*	*UB*	*p*	*d*
Harmfulness					

		95% CI			
Comparison	M_{diff}	*LB*	*UB*	*p*	*d*
Purity - Harm	-2.15	-2.60	-1.71	< .01	2.95
Purity - Disgusting Non-Moral	0.85	0.65	1.05	< .01	0.91
Harm - Disgusting Non-Moral	3.00	2.60	3.40	< .01	4.46
Immorality					
Purity - Harm	-0.52	-0.90	-0.14	< .01	0.58
Purity - Disgusting Non-Moral	1.11	0.84	1.38	< .01	1.10
Harm - Disgusting Non-Moral	1.63	1.27	1.99	< .01	1.75
Weirdness					
Purity - Harm	0.50	0.22	0.79	< .01	0.60
Purity - Disgusting Non-Moral	-0.21	-0.51	0.09	.16	0.26
Harm - Disgusting Non-Moral	-0.71	-1.03	-0.40	< .01	0.83
Disgust					
Purity - Harm	-0.07	-0.51	0.36	.74	0.07
Purity - Disgusting Non-Moral	-0.54	-0.88	-0.20	< .01	0.58
Harm - Disgusting Non-Moral	-0.47	-0.90	-0.04	.03	0.51

Note. M_{diff} = Mean difference, in the direction indicated by the Comparison column. CI = Confidence Interval, with *LB* and *UB* indicating the lower and upper bounds, respectively. *p* = p-value from a paired samples t-tests, which were used because the same participants provided measurements for harmfulness, immorality, weirdness, and disgust. *d* = Cohen's *d* measure of effect size difference.

Response Time Data

The mean response times of the retained behavior set, measured in seconds until first click, ranged from 3.80 to 11.40 seconds. As can be seen, the standard deviations were quite high (range: 1.44 – 11.70; behavior-level breakdown of response times can be found in Appendix C). This was probably due to participants taking breaks during the pilot study, which they were encouraged to do given the study took 45 minutes on average to complete. Because of the negatively skewed distribution, I considered the median as the more informative metric of response time (average Mdn = 4.63, range: 3 – 9).

Moral Judgments Under Time Pressure

The goal of the main experiment was to test whether time pressure would change how frequently people indicate a purity behavior is immoral because it is harmful vs disgusting. As in the Pilot Study, participants in this experiment responded with one of two reasons why they think a behavior is immoral – “has a victim” or “is disgusting”. Participants made their choice under two time-pressure conditions: timed and untimed. In the timed condition, participants had a limited time to make their choice, depriving them of cognitive resources that would be necessary to think of possible harmed victims, if indeed such victims are post-hoc rationalizations. In the untimed condition, participants had unlimited time to make their choice, leaving them with ample cognitive resources. The time pressure manipulation allows a test of the competing predictions of MFT and TDM, as described in the introduction.

Method

Participants

Participants were 150 Brooklyn College undergraduate students recruited from the Brooklyn College and Baruch College institutional research subject pools. Students participated in exchange for course credit. Outliers were defined as any participant showing no variance in forced choice responses (that is, responded with only “has a victim” or “is disgusting” for all behaviors across behavior types and time pressure conditions). Furthermore, any participant with greater than 25% missing data was excluded from all analyses. The mean age of the final sample was 21 years ($SD = 3.9$) and was 66.7% female, 29.7% Hispanic, 24.1% Asian, 6.9% black, 19.3% white, 20.0% multiple races/unknown.

Study Design and Procedure

This study followed a 2 (timed or untimed) × 3 (Moral judgments of purity, harm, and disgusting non-moral behaviors) fully within-subjects design. All participants responded to all behavior types under both timed and untimed conditions, the order of which was fully randomized. To identify an appropriate time limit for the timed condition, participants The study was programmed in Psychopy (Pierce et al., 2019).

Administration. Participants began with a timer calibration round where they completed an untimed moral judgment task. To allow for a manipulation check, participants also completed an untimed math task. The initial moral judgment and math tasks were presented in random order. As detailed below, the response latencies on these tasks were used to determine the time limit for subsequent timed conditions.

Following the timer calibration round, participants completed a timed math task as a manipulation check on this approach to setting time limits. Next, participants completed the untimed and timed moral judgment tasks in random order; data from these blocks were used to test the research hypotheses. Lastly, participants judged the immorality of all behaviors, and finally provided demographic information. I describe each task in more detail in the following sections, in the order seen in Figure .

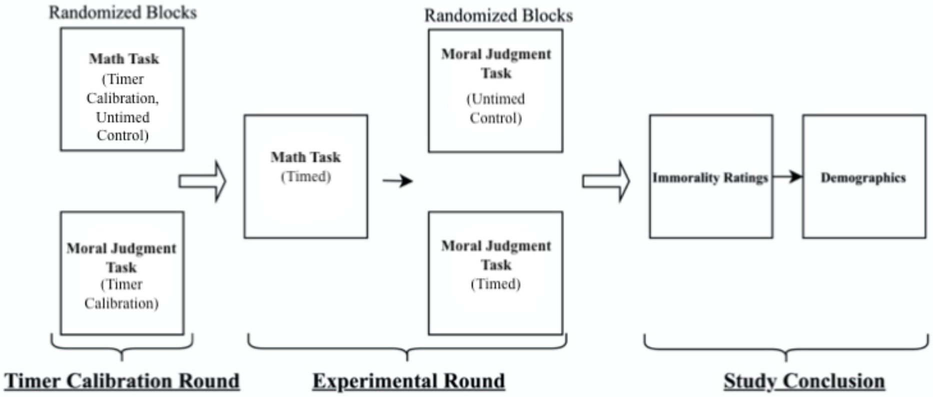

Randomized Blocks
Math Task (Timer Calibration, Untimed Control)
Moral Judgment Task (Timer Calibration)
Math Task (Timed)
Randomized Blocks
Moral Judgment Task (Untimed Control)
Moral Judgment Task (Timed)
Immorality Ratings
Demographics
Timer Calibration Round
Experimental Round
Study Conclusion

Figure 2. Overview of the study procedure.

Timer Calibration Round[2]

Moral Judgment Task (Untimed). Given that participants may vary in their reading speed, the experiment determined the time limit to be used in subsequent timed blocks individually for each participant, instead of using the same time limit for all participants. The purpose of the moral judgment task presented in the timer calibration round was determine the time limit to be used for each participant, based on their individual response time under untimed conditions. In this round, participants were randomly presented with 5 stimuli of each behavior type (harm, purity, disgusting non-moral) for a total of 15 behaviors. These behaviors were the left-over behaviors that were excluded from the retained behavior set through piloting. The data of interest for this task was the response time across the 15 behaviors; moral judgments were not analyzed further. Specifically, the timer for the timed moral judgment task in the experimental

[2] For a detailed explanation of how participant countdown times were calculated, please see Appendix D.

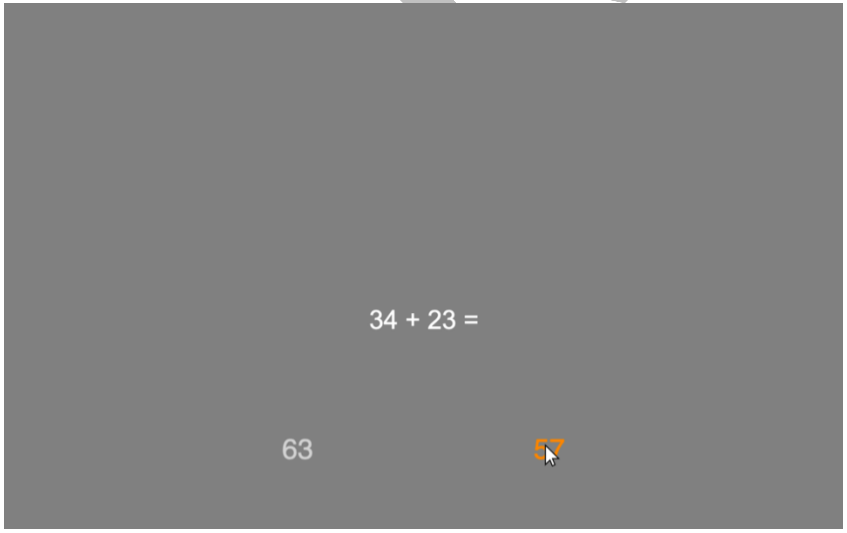
round was calculated for each participant i using the following formula: Timer Length[3] $= Mdn_{\mathrm{i}} - (0.25 \times MAD_{\mathrm{i}})$, where Mdn_i is the participant median response time and MAD_{i} is the participant response time median absolute deviation. This approach allows participants to have their own baselines while still receiving a standardized manipulation.

The moral judgment task in the timer calibration round was largely identical to the moral judgment task previously described in the pilot study, as were the response options. Briefly, participants were presented with short, one-sentence descriptions of harm, purity, or disgusting non-moral behaviors, indicating why they thought each behavior was immoral by selecting a harm option or disgust option. The only departures from the pilot moral judgment task was that here, the instructions encouraged participants to take as much time as they wanted in deciding why they thought the behaviors were immoral. The other departure was that the placement of the harm and disgust response options on the computer screen was counterbalanced. For half of the participants, the harm option appeared on the left side of the screen, and for the other half, it appeared on the right side of the screen. This was to control for potential effects handedness.

[3] I opted for the median and median absolute deviation because they are more robust to outliers than the more common mean and standard deviation. If a participant were to take abnormally long to respond to a trial, it is less likely that the measure of variability (the absolute deviation) would exceed the average (the median), relative to the mean and standard deviation. If e.g. the standard deviation exceeds the mean, the time countdown would be equal to a negative number, causing the experiment to crash.

Math Task (Untimed). A key assumption of this study is that time pressure limits cognitive resources needed for effortful reasoning. To provide a check on this assumption, participants responded to math problems under both timed and untimed conditions. If effortful reasoning is limited by time pressure, then participants should choose the correct response less frequently in the timed condition compared to the untimed condition. Previous research supports the use of math accuracy as a test of whether a manipulation has successfully reduced cognitive resources (e.g., Deck & Jahedi, 2015; Drichoutis & Nayga, 2017). Gray and colleagues' (2014) time pressure study used a similar strategy. While the experiment calculated the timers for moral judgment and math tasks separately, it applied the same formula to both tasks (MDN_i – [MAD$_\text{i}$ x 0.25]). Thus, if time pressure affects performance on the math task, it should also affect performance on the moral judgment task.

Figure 3. Example screenshot of an untimed trial from the math task.



Participants completed the untimed math task in the timer calibration round. Participants were presented a random selection of 7 out of 14 possible addition and subtraction problems. They responded by choosing one of two possible responses, with the correct response appearing on the left or right side of the screen in a random fashion. Participants were instructed to take as much as they wanted when deciding the correct response. A screenshot of the untimed math task from the timer calibration round is presented in

Figure . The data of interest for this task are the mean response time across the 7 math problems.

***Experimental Round*.**

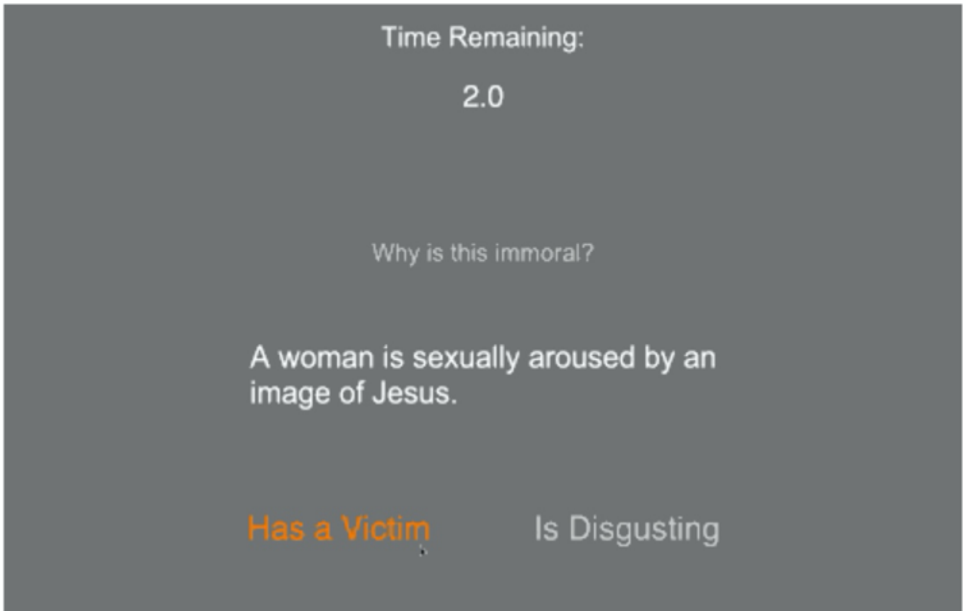
Math Task (Timed). Because solving math problems might be cognitively taxing in itself, irrespective of time pressure, I intentionally presented the timed math task at the beginning of the experimental round for all participants. This ensured participants were equally taxed by the time they made their moral judgments. Prior to the timed math task, participants were given the following instructions:

> "Beware: THIS TIME you will have to answer before a timer is over! That is, your task will be to click the correct answer before the timer reaches zero (the timer will be at the top of the screen).
>
> Decide as quickly as you can – use your gut!
>
> Be ready – as soon as you leave this page, the timer will start!"

After reading these instructions, participants clicked to advance to the next page, where they began the math trials. The next page, which was presented before every timed math trial, was a screen with the phrase "get ready…", which lasted for 2.5 seconds. Next, participants were presented a math trial, which was identical to

Figure except with a digital countdown at the top center of the page indicating the remaining time in the trial. As described above, the time limit was a function of participants' response time during the untimed math task. Participants either selected the response they thought was correct before the timer expired, or the timer expired and advanced them to the next page, whichever came first. Participants completed 7 trials, consisting of the 7 math problems that were not used in the untimed math task from the timer calibration round.

Moral Judgment Tasks. After the timed math task, participants completed timed and untimed moral judgment tasks in randomized order. The untimed moral judgment task was identical to the moral judgment task presented in the timer calibration round[4]. For the timed moral judgment task, participants received the same instructions as the timed math task, except the instructions were in reference to moral judgments instead of math. In other words, participants were told that they needed to respond before the timer finished, and to respond quickly. As with the timed math task, participants were presented with a screen showing the phrase “get ready…”, which lasted for 2.5 seconds, prior to every timed moral judgment trial.



Time Remaining:

2.0

Why is this immoral?

A woman is sexually aroused by an image of Jesus.

Figure 4. Example trial from the timed moral judgment task.

[4] While I could have used the moral judgment task from the timer calibration round as the control condition, as with the untimed math task, I opted for a separate untimed control condition to ensure that participants were uniformly cognitively taxed by the time they were given the untimed moral judgments.

Each trial showed a countdown timer at the top center of the screen indicating the remaining time for the trial (see Figure for a screenshot of a timed moral judgment trial). As described above, the time limit was a function of participants' median response time during the untimed moral judgment task that was presented during the timer calibration round.

In both the timed and untimed moral judgment tasks, participants responded to 15 behaviors, 5 of each type (harm, purity, disgusting non-moral). These behaviors were presented via a balanced random sample from the pool of 30 behaviors retained from the pilot study. More specifically, at the within-participant level, the task presented a unique set of 15 behaviors in the timed condition and a different unique set in the untimed condition, with behaviors presented in a fully randomized fashion (i.e. the presentation of behavior types, as well as behaviors within each type, was completely random). At the between-participant level, presentation of individual behaviors in timed vs untimed conditions was random, meaning that for one participant, "someone has sex with a chicken before eating it" might have appeared in the timed condition, while for the another participants, this behavior might have appeared in the untimed condition.

Study Conclusion. After completing the moral judgment tasks, participants rated the immorality of all behaviors in random order. The prompt read "how immoral is this behavior?", and responses were on a 7-point Likert scale where 1 = *Not at all immoral* and 7 = *extremely immoral*. I also collected ratings of political ideology, where participants responded to the statement "when it comes to politics, I see myself as…" on a 7-point Likert scale where 1 = *very conservative* and 7 = *very liberal*. Immorality ratings and political ideology were collected for exploratory purposes only and are not included in the main analyses. Participants also provided basic demographics of racial and gender identity, ethnicity, age, religious affiliation, state of birth (if born in the US), and education in years. These demographics were collected for sample reporting and exploratory purposes only and are not included in the main analyses.

Results

Manipulation Check: Timed vs. Untimed Math Task

Consistent with Gray et al. (2014), I used performance on a math task to demonstrate that time pressure depletes cognitive resources, which should limit participants' ability to engage in effortful reasoning. To ensure that the time pressure manipulation limited participants' cognitive resources, I conducted a mixed effects logistic regression with correlated random intercepts for participants and stimuli, with time pressure (timed/untimed) as the fixed effect independent variable and math response (correct or incorrect) as the dependent variable. Trials with no response were treated as incorrect responses. The timed condition was strongly associated with greater odds of an incorrect response (Log Odds = -6.70, 95% CI [-7.37, -6.03], < .001[5], *OR* = 0.00, Pseudo-R^2 [fixed effects] = 0.58, Pseudo-R^2 [full model] = 0.83). In an otherwise identical analysis where I only included trials with responses - that is, I excluded trials where the participants were not able to respond before the end of the timer - the timed condition was again associated with greater odds of an incorrect response (Log Odds = -6.92, 95% CI [-7.64, -6.21], < .001, *OR* = 0.00, Pseudo-R^2 [fixed effects] = 0.50, Pseudo-R^2 [full model] = 0.86). Participants were considerably less likely to give the correct response to math questions under time pressure, indicating time pressure successfully depleted cognitive resources.

Time Pressure and Moral Judgments

Exploratory Analyses. An exploration of the data using Fisher's exact tests showed that, irrespective of time pressure condition, participants overall found the harmful acts the most

[5] P-values for this analysis, as well as the analysis reported in Table 2, were calculated with using the Kenward and Roger (1997) method.

harmful, followed by purity, followed by disgusting non-moral acts. More specifically, participants responded with the harm option more frequently in the harm condition (74.0%) compared to the purity condition (27.2%; *OR* = 7.69, 95% CI [6.67, 9.01], *p* < .001) and disgusting non-moral condition (11.8%; *OR* = 23.08, 95% CI [18.88, 28.31]). Participants also responded with the harm option more frequently in the purity condition than the disgusting non-moral condition (*OR* = 2.99, 95% CI [2.46, 3.66]), indicating that participants found purity behaviors more harmful than disgusting non-moral behaviors.

Table shows the results of Fisher's exact tests between time pressure conditions (timed vs. untimed), both overall (i.e. irrespective of behavior type) and within behavior type conditions. Ignoring behavior type, participants overall responded with the harm option at similar frequencies in the untimed condition compared to the timed condition. As for within-behavior type comparisons, MFT predicts that participants should choose the harm option more frequently in the untimed purity condition relative to the timed purity condition. These preliminary results show that participants responded with the harm option at similar frequencies in the timed and untimed conditions, providing some preliminary evidence against MFT. As for the harm condition, both MFT and TDM predict that time pressure should have no influence over how frequently people find harmful behaviors immoral because they harmful. However, participants responded with the harm option more frequently in the untimed condition compared to the timed condition. That is, participants saw harm behaviors as less harmful when responding under time pressure. This finding contradicts both MFT and TDM. As for the disgusting non-moral condition, TDM predicts that participants should choose the harm option at similar frequencies in the untimed condition as in the timed condition. Indeed, participants responded with the harm option at similar frequencies in the untimed and timed disgusting non-moral conditions, providing preliminary evidence for TDM.

Table 3.

Frequencies (in Percentage) of the Harm Option in the Moral Judgment Task with Fisher's Exact Tests

	Percentage				95% CIs	
Behavior Type	Untimed	Timed	*p*	*OR*	Lower	Upper
Harm	78.40	72.91	.01	0.74	0.58	0.95
Purity	28.27	29.34	.68	1.05	0.83	1.34
Disgusting Non-Moral	12.27	11.50	.69	0.93	0.67	1.29
Total	39.64	37.82	.23	0.93	0.82	1.05

Note. OR = odds ratio, *p* = *p*-value, CIs = confidence intervals.

Main Analysis. Increasingly, it has been recommended that researchers control for variation between stimuli, as doing so has been shown to alter the effect sizes found in canonical social psychological data sets by as much as 60% (Judd, Westfall, & Kenny, 2012). To more stringently test the competing hypotheses of MFT and TDM, I controlled for participant- and stimulus-level variation using mixed effects logistic regression with behavior type, time pressure, and behavior type × time pressure interaction terms as fixed effects, with correlated random intercepts for participants and for individual behaviors (i.e. individual stimuli within each level of behavior type).

Table 4.

Mixed Effects Logistic Regression Predicting Frequency of Harm Option in Moral Judgment Task

Fixed Effects	95% Confidence Interval

Predictor	Est.	2.5%	97.5%	*OR*
(Intercept)	-2.23***	-2.61	-1.84	0.11
Harm	3.71***	3.20	4.21	40.85
Purity	1.13***	0.64	1.62	3.10
Timed	-0.09	-0.41	0.23	0.91
Harm × Timed	-0.25	-0.66	0.16	0.78
Purity × Timed	0.17	-0.24	0.57	1.19

Random Effects

Grouping Variable	Predictor	*SD*
Participant	(Intercept)	0.77
Why Immoral Stims	(Intercept)	0.46

Grouping Variable	*N*	*ICC*
Participant	150	0.15
Why Immoral Stims	30	0.05

Note. SD = standard deviation, which quantifies and standardizes the raw variation in the harm option attributable to the given grouping variable. *ICC* = intraclass correlation coefficient. The grouping variable denotes which variables were modeled with random intercepts. *Harm* and *Purity* denote behavior type conditions. Tests were two-sided, alpha = .05.

Table 4 reports results from this mixed effects model, for which the dependent variable is frequency of selection of the harm option. The intercept, or reference group, is the untimed disgusting non-moral condition. Marginal effects of harm and purity behavior types, as well as the marginal effect of time pressure, are reported along with interaction terms. Holding the marginal effect of purity, time pressure, and interaction terms constant (as well as accounting for the unique variance associated with participants and behaviors), the marginal effect of harm indicates that participants more frequently responded with the harm option in the harm condition, irrespective of time pressure condition, than they did in the untimed disgusting non-moral condition. This was also true for purity behavior type, though with a much smaller effect size (as indicated by the odds ratio). The marginal effect of time pressure, denoted by the *timed* term, shows a near-zero effect size.

MFT posits that harm concerns, in the case of purity, represent post-hoc rationalizations stemming from effortful reasoning intended to persuade others. As such, MFT predicts that in the purity condition, participants should choose the harm option more frequently in the untimed condition compared to the timed condition. The theoretically-relevant term then, is the purity × time pressure interaction, which tests whether frequency of the harm option within the purity condition varies as a function of time pressure, while simultaneously holding constant the unique contribution of harm and purity behavior conditions (irrespective of time pressure), time pressure condition (irrespective of behavior type), and a harm × timed interaction. It also holds constant the random-effects terms modeling between-subject and between-behavior variation that may also affect the frequency of the harm option. The effect of the purity × time pressure interaction was relatively small, and not statistically significant, indicating evidence against MFT.

The harm × timed interaction tests whether the frequency of the harm option varies by time pressure condition, within the harm behavior type. While the coefficient was same direction of the exploratory analysis, which showed a statistically significant difference such that participants responded with the harm option more frequently in the untimed harm condition, the harm x time pressure interaction did not emerge statistically significant. That is, while this analysis does show that participants more frequently responded with the harm option in the untimed harm condition, the effect was too weak relative to the noise in the study to be considered meaningful. This result would consistent with both MFT and TDM, each of which predict that harm judgments are driven by harm concerns and that time pressure should not affect the frequency of the harm option for harm judgments. Figure 5 shows an interaction plot visualizing the results. In short, if using the stricter mixed effects logistic regression approach, I found that the unique combinations of time pressure and behavior type, be it harm or purity, did not bear meaningful relationships with the frequency of the harm option in either purity or harm behavior conditions. This provides evidence against MFT, which predicts that participants should have responded with the harm option more frequently in the untimed purity condition than in the timed condition. This also provides some evidence for TDM, which allows for no difference between untimed and timed purity conditions, and also predicts the finding that participants did not choose the harm option less frequently in untimed disgusting non-moral condition compared the timed condition.

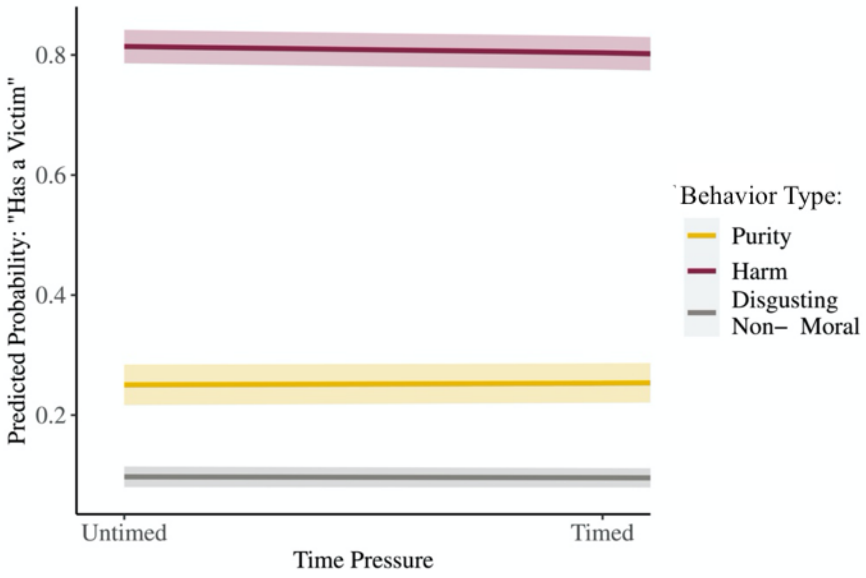


Figure 5. Predicted probabilities plot visualizing results from the mixed effects model. Shading indicates 95% confidence intervals.

Post-Hoc Power Analysis

Because I was only able to collect data from 150 participants, as opposed to the preregistered 200 participants, I conducted a power analysis using Monte Carlo simulations ($n = 1{,}000$) to assess the study's power to detect the theoretically-critical time pressure × behavior type interaction. To identify a non-trivial, though realistic, effect size, I used the upper bound of the confidence interval for the purity by time pressure interaction term (log odds = 0.57, $OR = 1.77$). This assessment, which assumes that the true effect is at least as high as the upper bound of the current confidence interval ($OR = 1.77$), indicated 75.10% power to detect an effect of this size, 95% CIs: [72.30, 77.75]. This indicates that the power of the study is currently below the convention of 80% power to detect an effect of interest, though see Zhang et al. (2019) for a discussion of the limitations of retrospective estimations of statistical power.

Discussion

The goal of this research was to test competing theoretical claims of MFT and TDM about the role of harm-based reasoning in moral judgments about harmless purity behaviors. MFT argues that harm concerns expressed toward purity behaviors represent post-hoc rationalizations. Therefore, MFT predicts that, under time pressure, people should be less likely to express concerns about harm toward purity behaviors, as compared to when they have ample time. TDM, on the other hand, argues that harm is necessary for people to see behaviors as immoral. Therefore, TDM predicts that time pressure should either make people more likely, or equally likely, to express concerns about harm toward purity behaviors, as compared to ample time. Contrary to MFT, results revealed no effect of time pressure on judgments about purity behaviors. The successful manipulation check suggests that the lack of an effect of time pressure on judgments about purity behaviors is not because the time pressure manipulation was not depleting. Specifically, participants performed worse on a math task when under time pressure compared to an untimed condition. The same formula was used to determine the time limit for responding to the moral behaviors, so it is likely that participants were depleted in the moral judgment task as well.

The finding that time pressure did not affect judgments about purity behaviors provides support for TDM’s hypothesis that time pressure either increases, or does not affect, the extent to which people see harm in purity behaviors. This provides evidence for TDM’s claim that intuitive concerns about harm cause moral judgments. One way that results from this study differ from Gray et al. (2014) is that these authors found that time pressure causes people to see more harm in purity behaviors as compared to an untimed condition, whereas this study found no effect of time pressure on judgments about purity behaviors. By my view, Gray’s finding would make more sense than ours, as having time to think should allow people to recognize the lack of harm and alter their response accordingly. It is not clear why harm judgments were equivalent in the timed and untimed conditions in this study.

Furthermore, I unexpectedly found that for the harm condition, time pressure was associated with fewer harm concerns and greater disgust concerns. That is, due to the forced-choice nature of this study, a lower frequency of harm concerns necessarily indicates higher frequency of disgust concerns. This finding contradicts both MFT and TDM, each of which propose that condemnation of explicitly harmful behaves is driven by concern about harm, not disgust. However, the finding that time pressure was associated with a lower frequency of the harm option (and therefore greater frequency of the disgust option, given the forced-choice design) is somewhat consistent with some previous research suggesting that disgust may be more important to judgments of harm behaviors than previously thought (Giner-Sorolla & Chapman, 2017). Indeed, disgust predicts condemnation of not just purity behaviors, but harmful behaviors as well (Chapman & Anderson, 2014). In fact, one perspective seeking to explain these findings argues that disgust is predominately related to judgments of moral character rather than judgments of behaviors, such that harmful behaviors may elicit disgust insofar as they indicate bad moral character (Chapman, 2018). It could be that participants in this study inferred worse moral character from the harm behaviors in this study than the purity behaviors, especially since the pilot data showed that the retained harm behaviors were rated as more immoral on average than the purity behaviors. Importantly, though, the evidence for this was mixed, as there were no statistically significant differences in harm concerns in the stricter tests. Furthermore, at least one previous finding indicates that purity behaviors drive stronger character inferences relative to harm behaviors, even though the purity behaviors themselves were considered less immoral (Uhlmann & Zhu, 2014). Ultimately, more research is needed to understand when character inferences are made, which kinds of behaviors elicit them, and what the role of effortful reasoning might be in character-based judgments.

Limitations

While the lack of an effect of time pressure on judgments about purity behaviors may indicate support for TDM, one limitation of forced-choice responses is the loss of resolution in measurement. Rather than indicating degree of agreement, participants are forced to wholly endorse one concern over the other. This could reduce statistical power, which when compounded by the premature ceasing of data collection due to the COVID-19 pandemic, may be responsible for null findings. Furthermore, there is some risk in concluding too much from null findings as they cannot distinguish between a “real” absence of a proposed phenomenon and an absence induced by study design or measurement.

One other limitation of this study could be that the time pressure manipulation was unsuccessful at depleting cognitive resources. While the manipulation check did show that participants were less accurate at math when responding under time pressure, it could be that the constraints of time pressure have different effects on math and moral judgments. For example, moral judgments might require greater time pressure than math to affect the effortful reasoning involved in making moral judgments. Alternatively, moral judgments might be more sensitive to time pressure than math, encouraging guessing during the moral judgment task. Another possibility is that the processes implicated in moral judgment-based effortful reasoning are more complex than, or incommensurable to, effortful reasoning that drives math accuracy. If true, moral judgments would require a different approach for limiting effortful reasoning.

Implications and Future Directions

Much research in moral psychology has focused on contrasting harm with purity and identifying their respective drivers. The current research is unfortunately limited in its ability to reveal much about these processes due to the null findings. I did identify, albeit provisionally, some evidence linking disgust to harm. While this research failed to find evidence supporting Moral Foundations Theory’s claim that harm concerns sometimes constitute a post-hoc rationalization, there has been surprisingly little empirical attention devoted to mapping out the causal order of harm concerns, effortful reasoning, and moral judgments of right and wrong.

While time pressure is one way to test these competing hypotheses, it may not be the best approach. Although time pressure may have an influence over the ability to deploy effortful reasoning, it also limits reading time, and causes other unintended side effects such as feelings of urgency and increased arousal (Svenson & Maule, 1993). Future research could try other, more direct strategies for impeding effortful reasoning, such as through cognitive load. Some research has found that people are less concerned with purity when they are cognitively depleted (Wright & Baril, 2011), though this finding is at odds with other work showing that people are more concerned with harm after activating analytic (Yilmaz & Saribay, 2017) and abstract thinking

(Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014). It could be that harm concerns are higher *both* when under high cognitive load and when analytic or abstract thinking is activated, with purity concerns arising only when there is a relatively moderate amount of cognitive resources available. That is, a U-shaped curve where people care most about harm when they are cognitive resources limited or amplified, with purity arising in the in-between levels of cognitive resources. Or perhaps, more simply, cognitive load has the unintended effect of causing greater cognitive engagement.

Ultimately, much remains to be learned about the role of effortful reasoning in moral judgments. While these issues are theoretically important, they may also be culturally timely. With the ongoing COVID-19 pandemic, which to date has infected over 6,000,000 people and killed nearly 200,000 (*US Historical Data*, 2020), the role of effortful reasoning may have never been more important. Moral judgments have been found to affect adherence to CDC guidelines (Everett, Colombatto, Chituc, Brady, & Crockett, 2020), and effortful reasoning has been shown to predict accurate beliefs about COVID-19, stronger even than political ideology (Pennycook, McPhetres, Bago, & Rand, 2020). Taken together, the role of effortful reasoning in moral judgments, particularly as they pertain to adherence to COVID-19 safety protocols, are important not just in theory; they could potentially save lives if robust findings are effectively incorporated into policy and public health messaging. I look forward to seeing where future research on these topics takes the field.

References

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*(3), 439–449.

Chapman, H. A. (2018). A component process model of disgust, anger, and moral judgment. *Atlas of moral psychology*, *70*.

Chapman, H. A., & Anderson, A. K. (2014). Trait physical disgust is related to moral judgments outside of the purity domain. *Emotion*, *14*(2), 341–348. https://doi.org/10.1037/a0035120

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.

Clifford, S., & Wendell, D. G. (2016). How disgust influences health purity attitudes. *Political Behavior*, *38*(1), 155–178.

DeScioli, P., Gilbert, S. S., & Kurzban, R. (2012). Indelible victims and persistent punishers in moral cognition. *Psychological Inquiry*, *23*(2), 143–149. https://doi.org/10.1080/1047840X.2012.666199

Edland, A., & Svenson, O. (1993). Judgment and decision making under time pressure. In *Time pressure and stress in human judgment and decision making* (pp. 27–40). Springer.

Everett, J. A. C., Colombatto, C., Chituc, V., Brady, W. J., & Crockett, M. (2020). The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic. https://doi.org/10.31234/osf.io/9yqs8

Gilet, A.-L., & Jallais, C. (2011). Valence, arousal and word associations. *Cognition & Emotion*, *25*(4), 740–746. https://doi.org/10.1080/02699931.2010.500480

Graham, J. (2015). Explaining away differences in moral judgment: Comment on gray and keeney (2015). *Social Psychological and Personality Science*, *6*(8), 869–873. https://doi.org/10.1177/1948550615592242

Graham, J., Englander, Z., Morris, J., Hawkins, C., Haidt, J., & Nosek, B. (2012). Warning bell: Liberals implicitly respond to group morality before rejecting it explicitly. Retrieved from https://papers.ssrn.com/abstract=2071499

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Chapter two - moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

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.

Gray, K., & Graham, J. (Eds.). (2018). Atlas of moral psychology, 608.

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.

Gray, K., Schein, C., & Cameron, C. D. (2017). How to think about emotion and morality: Circles, not arrows. *Current Opinion in Psychology*, *17*, 41–46. https://doi.org/10.1016/j.copsyc.2017.06.011

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. https://doi.org/10.1037/a0036149

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological inquiry*, *23*(2), 101-124.

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, *23*(2), 206–215.

Gray, K., & Wegner, D. M. (2012). Morality takes two: Dyadic morality and mind perception. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil.* (pp. 109–127). Washington: American Psychological Association. https://doi.org/10.1037/13091-006

Haidt, J. (2008). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. In J. E. Adler, L. J. Rips, J. E. Adler (Ed), & L. J. Rips (Ed) (Eds.), *Reasoning: Studies of human inference and its foundations.* (pp. 1024–1052). New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9780511814273.055

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

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

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

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The Innate Mind*, *3*, 367–391.

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.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, *103*(1), 54.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601-625.

Kerstholt, J. (1994). The effect of time pressure on decision-making behaviour in a dynamic task environment. *Acta Psychologica*, *86*(1), 89–104.

Lienard, P. & Boyer, P. (2006). *"Whence collective rituals? A cultural selection model of ritualized behavior". American Anthropologist.* ***108*** (4): 824–827. doi:10.1525/aa.2006.108.4.814.

Meyer-Rochow, V. B. (2009). Food taboos: Their origins and purposes. *Journal of Ethnobiology and Ethnomedicine*, *5*(1), 18. https://doi.org/10.1186/1746-4269-5-18

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. Psychological bulletin, 135(2), 303.

Outten, H. R., Lee, T., & Lawrence, M. E. (2019). Heterosexual women's support for trans-inclusive bathroom legislation depends on the degree to which they perceive trans women as a threat. *Group Processes & Intergroup Relations*, *22*(8), 1094-1108.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). The role of analytic thinking in moral judgements and values. *Thinking & reasoning*, *20*(2), 188-214.

Pennycook, G., McPhetres, J., Bago, B., & Rand, D. G. (2020). Attitudes about COVID-19 in Canada, the UK, and the USA: A novel test of political polarization and motivated reasoning.

Rottman, J., Kelemen, D., & Young, L. (2014). Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition*, *130*(2), 217–226. https://doi.org/10.1016/j.cognition.2013.11.007

Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, *41*(8), 1147–1163.

Schein, C., & Gray, K. (2017). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*. https://doi.org/10.1177/1088868317698288

Schwabe, C. W. (1996). *Unmentionable cuisine* (4th ed.). Charlottesville, Virginia: The University Press of Virginia.

Scott, S. E., Inbar, Y., & Rozin, P. (2016). Evidence for absolute moral opposition to genetically modified food in the united states. *Perspectives on Psychological Science*, *11*(3), 315–324.

Seidel, A., & Prinz, J. (2013). Sound morality: Irritating and icky noises amplify judgments in divergent moral domains. *Cognition*, *127*(1), 1–5.

Suri, R., & Monroe, K. B. (2003). The effects of time constraints on consumers' judgments of prices and products. *Journal of Consumer Research*, *30*(1), 92–104.

Svenson, O., Edland, A., & Slovic, P. (1990). Choices and judgments of incompletely described decision alternatives under time pressure. *Acta Psychologica*, *75*(2), 153–169.

Svenson, O., & Maule, A. J. (1993). Concluding remarks. In *Time pressure and stress in human judgment and decision making* (pp. 323–329). Springer Science & Business Media.

Uhlmann, E. L., & Zhu, L. (2014). Acts, persons, and intuitions: Person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science*, *5*(3), 279-285.

US Historical Data. (2020, September 20). Retrieved September 21, 2020, from https://covidtracking.com/data/national

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

Wright, J. C., & Baril, G. (2011). The role of cognitive resources in determining our moral intuitions: Are we all liberals at heart? *Journal of Experimental Social Psychology*, *47*(5), 1007-1012.

Yilmaz, O., & Saribay, S. A. (2017). Activating analytic thinking enhances the value given to individualizing moral foundations. *Cognition*, *165*, 88-96.

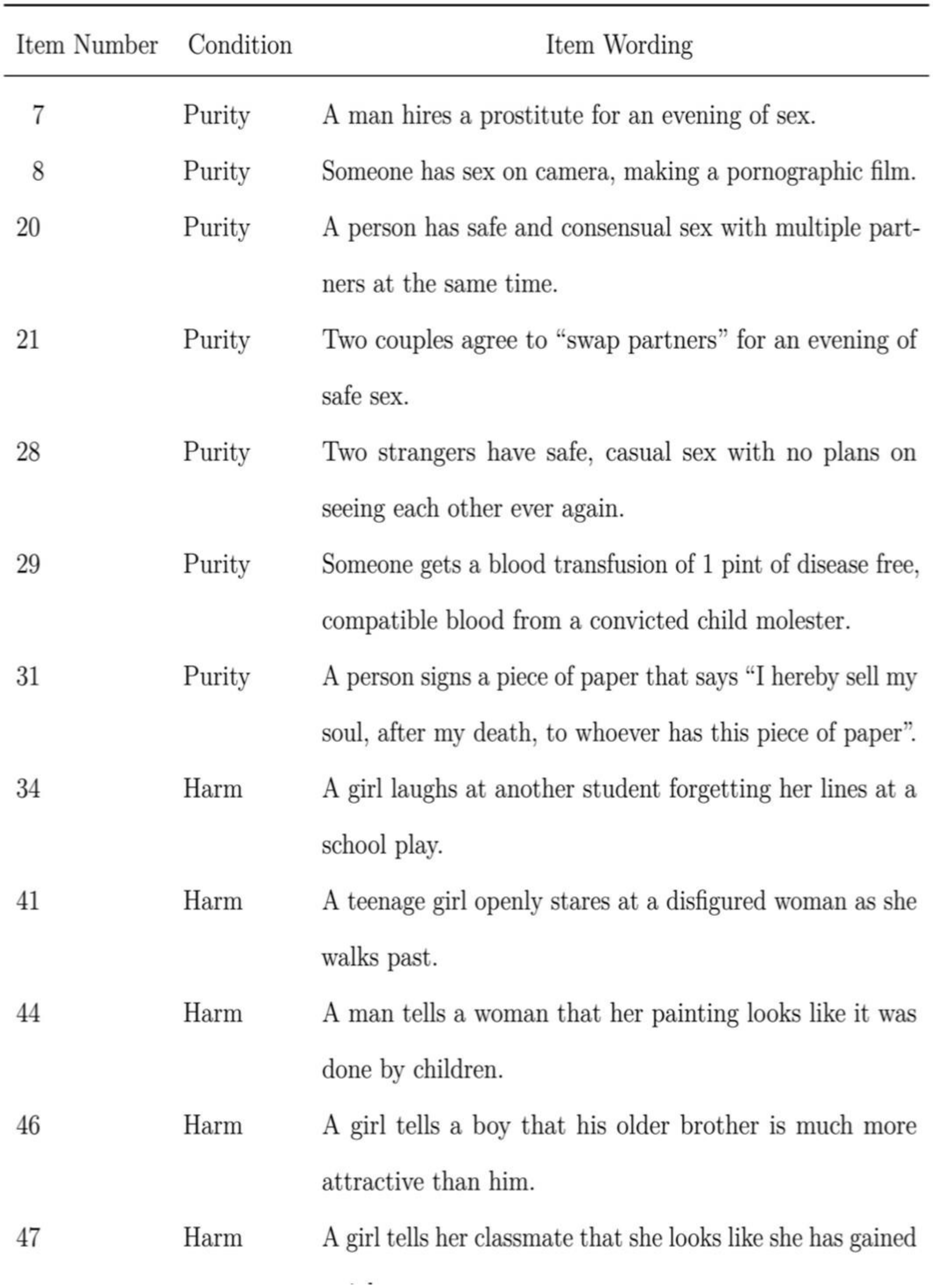
Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: Is it an informative and meaningful analysis? *General Psychiatry*, *32*(4), e100069–e100069. https://doi.org/10.1136/gpsych-2019-100069

Appendix A. Descriptive Statistics for Individual Behaviors from the Final Behavior Set

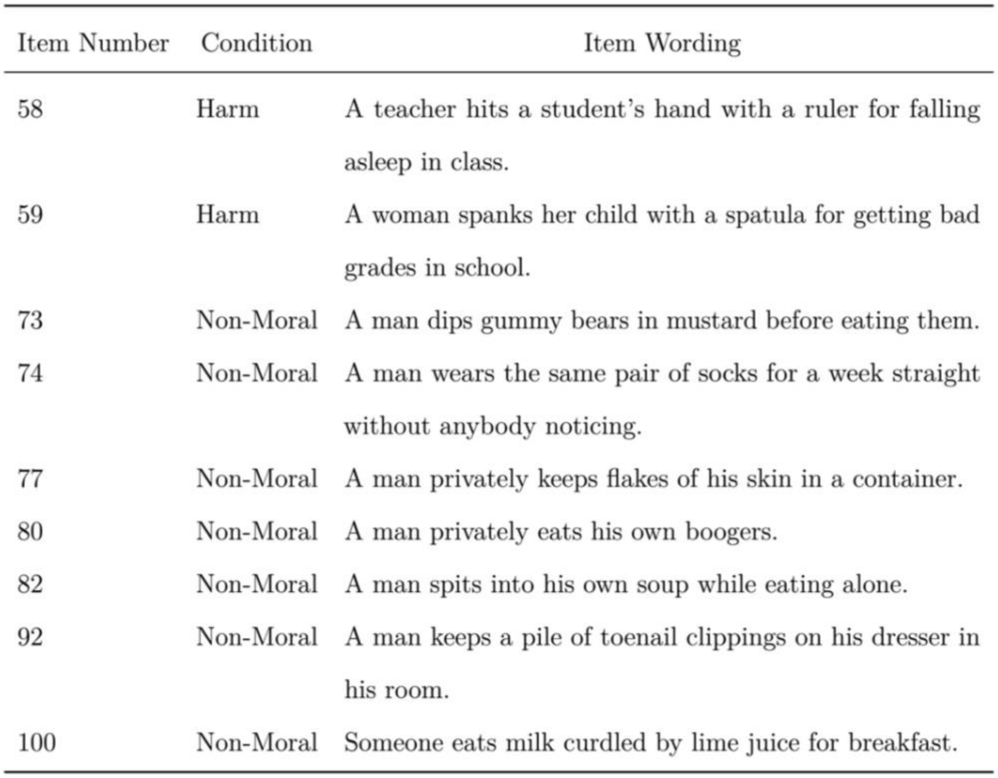
		Harm Option	Immorality		Harmfulness		Weirdness		Disgust	
Number	Behavior Type	%	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
7	Purity	0.48	2.36	1.29	2.72	1.46	2.68	1.34	3.04	1.31
8	Purity	0.36	2.48	1.45	2.00	1.19	2.20	1.41	2.28	1.57
20	Purity	0.36	2.60	1.50	2.28	1.31	2.40	1.53	2.36	1.44
21	Purity	0.36	2.68	1.38	2.12	1.24	3.00	1.32	2.84	1.57
28	Purity	0.48	1.96	1.24	1.84	1.07	1.64	1.00	1.92	1.29
29	Purity	0.40	2.00	1.50	2.08	1.26	2.27	1.48	2.15	1.52
31	Purity	0.44	2.81	1.42	2.62	1.50	3.85	1.22	2.54	1.63
34	Harm	0.88	3.16	1.38	4.32	0.69	1.88	1.20	3.08	1.38
41	Harm	0.92	2.68	1.11	4.20	0.82	2.44	1.23	2.68	1.31
44	Harm	0.88	2.54	1.50	3.96	1.08	2.08	1.32	2.58	1.47
46	Harm	0.96	2.60	1.32	3.92	0.91	1.72	0.84	2.36	1.38
47	Harm	1.00	2.80	1.23	4.28	0.68	2.12	0.97	2.84	1.38
58	Harm	0.88	3.16	1.25	4.52	0.65	3.04	1.24	2.64	1.29
59	Harm	0.88	2.88	1.37	4.42	0.81	2.23	1.21	2.42	1.39
73	Disgusting Non-Moral	0.08	1.46	1.03	1.31	0.84	3.73	1.22	3.42	1.30
74	Disgusting Non-Moral	0.12	1.84	1.46	1.52	1.16	3.20	1.35	3.60	1.19
77	Disgusting Non-Moral	0.08	1.96	1.54	1.77	1.18	4.04	1.22	4.04	1.11
80	Disgusting Non-Moral	0.04	1.85	1.54	1.54	1.21	3.54	1.30	3.81	1.17
82	Disgusting Non-Moral	0.08	1.72	1.34	1.64	1.15	3.64	1.32	3.96	1.21
92	Disgusting Non-Moral	0.08	1.92	1.50	1.68	1.34	3.84	1.21	3.80	1.16
100	Disgusting Non-Moral	0.08	1.64	1.35	1.68	1.22	3.60	1.12	3.64	1.04

		Harm Option	Immorality		Harmfulness		Weirdness		Disgust	
Number	Behavior Type	%	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Overall:		0.47	2.34	1.37	2.69	1.08	2.82	1.24	2.95	1.34

Appendix B. *Wording of Final Behavior Set*

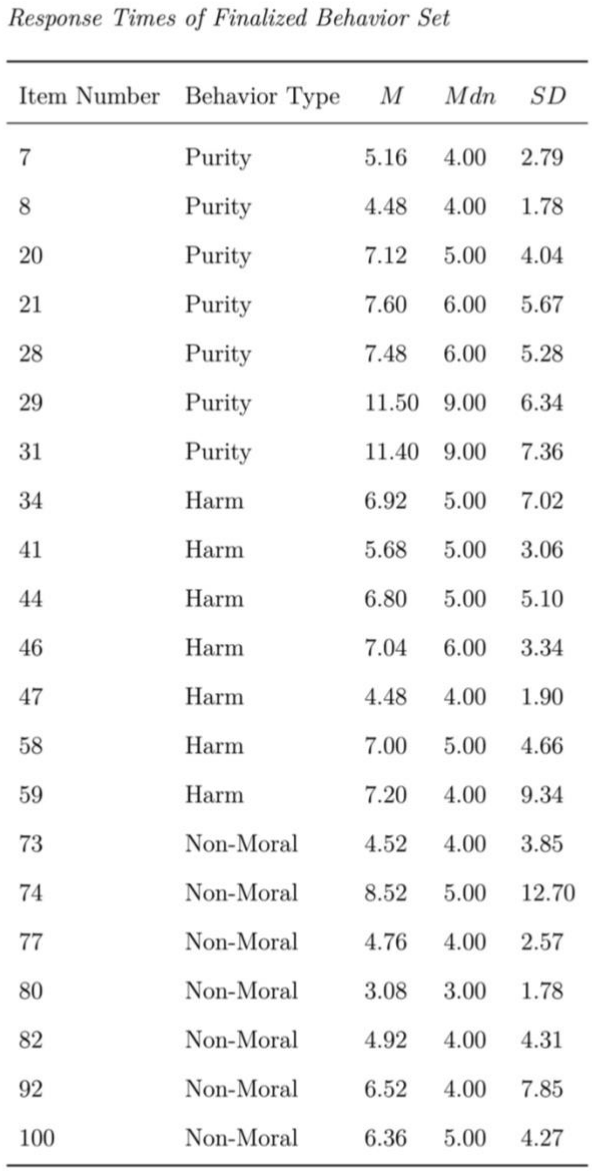


Item Number	Condition	Item Wording
7	Purity	A man hires a prostitute for an evening of sex.
8	Purity	Someone has sex on camera, making a pornographic film.
20	Purity	A person has safe and consensual sex with multiple partners at the same time.
21	Purity	Two couples agree to “swap partners” for an evening of safe sex.
28	Purity	Two strangers have safe, casual sex with no plans on seeing each other ever again.
29	Purity	Someone gets a blood transfusion of 1 pint of disease free, compatible blood from a convicted child molester.
31	Purity	A person signs a piece of paper that says “I hereby sell my soul, after my death, to whoever has this piece of paper”.
34	Harm	A girl laughs at another student forgetting her lines at a school play.
41	Harm	A teenage girl openly stares at a disfigured woman as she walks past.
44	Harm	A man tells a woman that her painting looks like it was done by children.
46	Harm	A girl tells a boy that his older brother is much more attractive than him.
47	Harm	A girl tells her classmate that she looks like she has gained



Item Number	Condition	Item Wording
58	Harm	A teacher hits a student's hand with a ruler for falling asleep in class.
59	Harm	A woman spanks her child with a spatula for getting bad grades in school.
73	Non-Moral	A man dips gummy bears in mustard before eating them.
74	Non-Moral	A man wears the same pair of socks for a week straight without anybody noticing.
77	Non-Moral	A man privately keeps flakes of his skin in a container.
80	Non-Moral	A man privately eats his own boogers.
82	Non-Moral	A man spits into his own soup while eating alone.
92	Non-Moral	A man keeps a pile of toenail clippings on his dresser in his room.
100	Non-Moral	Someone eats milk curdled by lime juice for breakfast.

Appendix C



Response Times of Finalized Behavior Set

Item Number	Behavior Type	*M*	*Mdn*	*SD*
7	Purity	5.16	4.00	2.79
8	Purity	4.48	4.00	1.78
20	Purity	7.12	5.00	4.04
21	Purity	7.60	6.00	5.67
28	Purity	7.48	6.00	5.28
29	Purity	11.50	9.00	6.34
31	Purity	11.40	9.00	7.36
34	Harm	6.92	5.00	7.02
41	Harm	5.68	5.00	3.06
44	Harm	6.80	5.00	5.10
46	Harm	7.04	6.00	3.34
47	Harm	4.48	4.00	1.90
58	Harm	7.00	5.00	4.66
59	Harm	7.20	4.00	9.34
73	Non-Moral	4.52	4.00	3.85
74	Non-Moral	8.52	5.00	12.70
77	Non-Moral	4.76	4.00	2.57
80	Non-Moral	3.08	3.00	1.78
82	Non-Moral	4.92	4.00	4.31
92	Non-Moral	6.52	4.00	7.85
100	Non-Moral	6.36	5.00	4.27

Appendix D

Procedure for Determining Countdown Time

Participants responded to both math and moral judgment tasks under time pressure. However, participants may take longer to respond to math problems compared to moral judgments, and vice versa for others. Countdown timers were thus calculated using identical, but independent, procedures for math and moral judgment tasks. To obtain baseline response times, I administered a timer calibration round at the beginning of the study, where participants completed an untimed math task and an untimed moral judgment task, in random order. During the timer calibration round, the experiment software recorded two median response times and median absolute deviations[6] for each participant: one for their responses to the math task, another for their responses to the moral judgment task.

Why This Procedure?

In Gray et al. (2014)'s time pressure study, all participants in the timed condition had to make their judgments in under 7 seconds. However, a uniform timer is likely to induce error variance, as participants differ in reading speed, comprehension, and a host of other attributes that influence response time. That is, some participants will find it very easy to respond under 7 seconds, thus having higher-than-average cognitive resources with which to make their judgments; likewise, some participants will find it hard, and thus be more likely to "click for the sake of clicking" to avoid letting the timer expire. To eliminate some of the error variance caused by unmodeled individual differences in response times, I chose to calibrate response times so that participants experience time pressure relative to their own baseline response times.