Examining the Effect of Exposure to Social Harm on Moral Disapproval and Perceived Harm, and the Moderating Role of Moral Identity

Kateřina Vaštová

S4642317

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number 43

Supervisor: Dr. Maja Graso

Second evaluator: MSc. Roxana Bucur

In collaboration with: Lisette Abels, Laura Keijzer, Lenka Kudelska, Maike Müller-Kuckelberg and Cindy Oosterhuis

July 3, 2024

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned

Abstract

In this thesis I question if the individual exposure to the threat of social harm could influence the differences in interpretation of ambiguously harmful social interactions. Specifically, if it affects perceived harm and moral disapproval. I hypothesized that a) exposure would lead to higher perceived harm and moral disapproval b) higher level of internalized moral identity will strengthen the relationship between exposure, perceived harm and moral disapproval. I used an online vignette-based experiment, randomly assigning participants either to the experimental or control condition. In experimental condition they were exposed to the threat of social harm in the form of social safety campaign, and in control they were not exposed to any stimuli. Then I measured participants' evaluations of two unrelated vignettes, showing ambiguous social harm. One vignette was about social exclusion, while one was about inappropriate comments in the workplace. Differences in outcomes of vignette dealing with social exclusion were not significant between conditions. For vignette dealing with inappropriate comments, exposure to social safety condition was associated with higher moral disapproval (p = 0.03; d = .299). Internalized moral identity did not moderate the relationship. The results are inconclusive, but subject to limitations due to the measures and materials used.

Keywords: social harm exposure, vignettes, moral disapproval, harm perception, internalized moral identity

Examining the Effect of Exposure to Social Harm on Moral Disapproval and Perceived Harm, and the Moderating Role of Moral Identity

In 2021, Chris Pratt published an Instagram post devoted to his wife for her birthday, thanking her for giving him an amazing healthy daughter. However, what was most likely intended as a simple post of appreciation, quickly became something more. Instantly, The Internet has decided that it was a dig at his ex-wife, with whom he has a visually impaired son with a heart problem. Chris Pratt faced an extensive social media backslash and judgement from one fraction of people, and with the other fraction of people defending him, Twitter threads on the topic quickly created a proverbial war zone (Osifo, 2021). This story is a great example of how morality, thought to be an adaptive concept, can hinder our lives and polarise us just us easily (Ellemers et al., 2019). But why did one group of people interpret his vaguely ambiguous message negatively, and judge what he did as terrible and morally wrong, while others did not?

This is particularly necessary to understand in our increasingly polarized world, as differences in moral evaluations pit people against each other, and lead to negative consequences. Namely, when one sees something as moral while the other does not it can lead to lower cooperation (Skitka et al., 2005), fuel ostracization (Täuber, 2019) and increase polarization and segregation (Clifford, 2019; Kovacheff et al., 2018). Finding what can be behind the differences in moral evaluation is the first step that can help facilitate better cooperation and restore social cohesion.

To find why moral evaluations differ, it is important to understand what is behind them. One thing that can trigger moral evaluation of an action, is perceiving that the action is harmful to someone or something (Haidt & Joseph, 2008; Schein & Gray, 2016; Turiel, 1983). Perception of what, and how much, is something harmful, is vastly different for everyone, especially if the situation has no signs of physical harm. Harm in social interactions, hereafter

referred to as social harm, is particularly ambiguous, as it is up to us to decide if the social interaction is causing someone suffering (Haslam, 2016). For example people differ in opinions on harmfulness of racial microaggressions (Midgette & Mulvey, 2024). The differences in perception of harm, in such situation, would then lead to differences in moral evaluation (Schein & Gray, 2018).

One of the factors behind these differences in harm perception could be individual exposure to social harm related concepts. Social harm is a growingly hot topic in public discourse, with some people and organisations stressing its importance and the danger of small remarks, such as microaggressions (Haslam, 2016; Haslam et al., 2021; Walsh, 2020; Wheeler et al., 2019). However, the exposure to social harm related concepts, for example on social media, will differ. Often, focus of the general public discourse is on the threat and danger of social harm. The differences in exposure to the threat of social harm could remind us, and make us more aware, of the concept of social harm, and increase the tendency to interpret ambiguous social situations as harmful and morally disapprove of the behaviour (Bleske-Rechek et al., 2023).

In the present study I manipulate exposure to the threat of social harm to see whether it is associated with higher harm perception and negative moral evaluation, such as moral disapproval, of behaviour in ambiguous social situations. However, the reason for the differences in evaluations of ambiguous social harm, such as Chris Pratt's message or microaggressions, could also be individual differences. For example, it was shown that moral concepts are more accessible for people with high moral identity (Aquino et al., 2009; Baumert & Schmitt, 2009), and moral identity is associated with stronger reactions to perceived slight (Skarlicki & Rupp, 2010). I investigate internalised moral identity as a moderator, predicting high internalised moral identity will strengthen the relationship between exposure to the threat of social harm and negative evaluation of ambiguous situations.

In the following sections, I address why should exposure to the threat of social harm theoretically affect what people evaluate as harmful and morally wrong. I explain what a moral evaluation is, its connection to harm, how we interpret ambiguous situations and why exposure to threat of social harm might make some perceive ambiguous situations as more harmful and trigger negative moral evaluation. I draw on Theory of Dyadic Morality (TDM; Schein & Gray, 2018), accessibility and negativity bias. I also explain why this may further be affected by individual differences in moral identity.

Theoretical Foundation

What Makes us Evaluate Situation From a Moral Viewpoint? Harm

First, I will define the concept of moral versus conventional evaluation, and why do we decide to evaluate something morally versus conventionally. Moral evaluations can be defined as deciding on whether something is fundamentally right or wrong, and to what extent (Kovacheff et al., 2018; Skitka & Mullen, 2002). Compared to conventional evaluation of something being right or wrong, moral evaluations of right or wrong are seen as universally true, seen as above laws and authority, and considered as more serious (Huebner et al., 2010; Smetana et al., 2018). When someone's action triggers negative moral evaluation, it also leads to calls for harsher punishments than for actions that triggered negative conventional evaluations and induces the willingness to restore what is perceived as proper moral order by any means necessary (Skitka & Mullen, 2002).

Researchers claim that morality is mostly automatic, and our decision to evaluate something morally versus conventionally is based on foundations (Haidt, 2001; Haidt & Graham, 2007; Schein & Gray, 2018). In this context, foundation is a core variable, that, when we see it or feel it, leads to moral evaluation under some boundary conditions. A commonly proposed foundation is harm, which was proposed to play part by Turiel (1983), is included in

the five foundations named by Moral Foundation Theory (MFT; Haidt & Graham, 2007), and is seen as the only moral foundation by TDM (Schein & Gray, 2018).

Perceived Harm and Negative Moral Evaluation Go Hand in Hand: TDM

Before delving into the reasons why higher social harm could lead to higher perceived harm and moral evaluation in ambiguous situations, I will establish the link between perceived harm and negative moral evaluation itself. As mentioned, harm is universally seen as a trigger for moral judgement (Haidt & Graham, 2007; Schein & Gray, 2018). However, it is important to note here, that when I mention harm, I am talking about perceived harm, not about objective harm. It is about interpreting the situation as harmful, not whether it is objectively harmful (Schein & Gray, 2018).

But how does perception of harm lead to negative moral evaluation? Under certain conditions, it seems to be a mutually reinforcing loop (Schein & Gray, 2016). First of all, for the harm to be registered as a moral offence, it has to fulfil two conditions: it has to be a violation of a social norm, while creating a negative affect (Schein & Gray, 2018). If that is fulfilled, TDM presents harm and moral judgement as mutually reinforcing variables forming cyclical dyadic loop. As Schein and Gray (2016) put it "what seems harmful seems wrong, and what seems wrong seems more harmful, and what seems more harmful becomes more wrong, and so on." (p. 62). So seeing even just a speck of harm or wrongness can lead to increased perception of harm and moral judgement through moralization, creating a spiral of harm being "both the cause and the consequence of moral disagreement" (Schein & Gray, 2018, p. 51). Altogether this means that increase in perceived harm should go hand in hand with increase in negative moral judgement, as the amount of perceived harm is a crucial for how much we morally condemn the action or disapprove of it.

The Role of Knowledge in Interpreting Ambiguous Situations

When evaluating the amount of harm present, we first rely on automatic judgements, and our brain automatically looks at past experiences and knowledge that could be relevant to use for the evaluation (Hjeij & Vilks, 2023). According to the availability heuristic, being exposed to something increases its accessibility and makes it more likely that we will use this in our decision process or for interpretation of ambiguous situations (Tversky & Kahneman, 1973). For example, consumption of media that focuses on crimes is associated with greater worry about being a victim of crime, while people who consume media focusing on climate change worry more about climate change (Andersen et al., 2024)

The weight we put on the information increases if the previous experience or stimuli is negative or threatening, as our brains are automatically wired to pay more attention to such information (Ito et al., 1998; Mikhael et al., 2021). For example, when participants were asked to form sentences about hostile vs kind behaviour, it later skewed their evaluations of ambiguous behaviour towards hostile or kind respectively (Srull & Wyer, 1979). However, effects of hostility activation affected evaluation up to 24 hours after the experiment, while kindness activation did not (Srull & Wyer, 1979), showing the comparable strength of threatening vs positive stimuli. We also tend to pre-emptively judge ambiguous situations as negative or morally wrong (Hester et al., 2020), especially after being exposed to threat (Neta et al., 2017). It is also connected to increased perception of harm itself (Bellet et al., 2018)

Effect of Exposure to the Threat of Social Harm on Harm Perception and Moral Disapproval

Based on the theory and evidence above, any clues that suggest harm should automatically receive a lot of attention, as they have negative and threatening aspect. The negativity and threatening aspects will then increase the accessibility and make it especially likely that they will be used in subsequent interpretation and evaluation (Ito et al., 1998). This should increase the likelihood that ambiguously harmful social situations will be interpreted

as harmful, and produce negative moral evaluation through the dyadic loop predicted by TDM (Schein & Gray, 2018). This is supported by research done by Bleske-Rechek and colleagues (2023), who specifically focused on ambiguous harm in social situations and discovered that exposing participants to social harm prime made them evaluate ambiguous sentences as more harmful. Research by Neta and colleagues (2017) found that highlighting social harm makes us interpret ambiguous situations in a negative light, providing additional support. Therefore, my first hypothesis is as follows:

H1: Participants exposed to the threat of social harm will evaluate ambiguous social harms as more harmful and they will show greater moral disapproval, compared to the unexposed control group.

The Moderator: Internalised Moral Identity

Other than contextual and environmental factors, perceived harm and moral evaluation are also dependent on individual differences (Gray et al., 2012). One such difference, that suggests itself particularly due to its connection to moral evaluation, is moral identity. Differences in identities in general have been shown to influence moral judgements (Leavitt et al., 2012). Moral identity seems to facilitate strong automatic response to observed injustice or harm (Skarlicki & Rupp, 2010). Moral identity has been divided into two parts, symbolic moral identity, focused on presenting oneself as moral, and internalised moral identity (Aquino & Reed, 2002). Internalised moral identity reflects the extent to which being moral is important to oneself (Aquino & Reed, 2002; Lutz et al., 2022). As the extent to which moral identity shapes moral judgement depends on the accessibility of it in one's self-concept (Aquino et al., 2009), in this paper I consider internalised moral identity.

If one has high internalised moral identity, the accessibility of morality should be greater as it means being moral is more important, and higher accessibility of certain identity leads to higher use of the information (Leavitt et al., 2012). In the past, studies have showed

internalised moral identity to have moderating effects on moral disapproval after observing social incivility (Lin & Loi, 2021), and it seems to be associated with increase in perceived offence severity (Barclay et al., 2014). Moreover, being seen as immoral is undesirable, so people try to maintain their image as being moral in their self-perception (Ellemers et al., 2019). It may be reasonable to assume that if one has higher internalised moral identity, they will morally disapprove more to protect themselves from feeling immoral, to avoid cognitive dissonance. Harm perception should also be increased, as it is a proposed foundation of moral disapproval (Schein & Gray, 2018).

H2: High internalised moral identity will strengthen the predicted positive relationship between exposure to the threat of social harm, moral disapproval and perceived harm.

Methods

Procedure and Participants

This study is part of a bigger research framework that was agreed upon and carried out by six bachelor thesis students. We designed the questionnaire, and with help from the supervisor, submitted the questionnaire for ethics approval granted by the ethical committee of the Faculty of Behavioral and Social Sciences at RuG. The survey was administered via an online form on Qualtrics, with responses collected in English. Participants were recruited through personal networks, such as Facebook and LinkedIn, and through Prolific. On personal networks we posted the questionnaire link with a uniform message about the nature and purpose of the questionnaire. On Prolific, we used the funding allocated to the thesis group by the faculty. The questionnaire started with informed consent, and at the end the participants had the chance to rescind the consent. In total, we recruited 227 participants, out of which 157 was from personal networks and 70 through Prolific. 51 was removed by the thesis supervisor in preparatory cleaning due to unfinished answers or not giving consent. The data from each group was then combined by the supervisor to protect anonymity.

To ensure high quality of responses, three attention checks were included throughout the questionnaire, asking the participants to select a certain answer, such as "*Please select "Somewhat Agree"".* According to our supervisor's advice, I only used answers that passed at least two out of three attention checks, meaning I further removed 15 participants. Out of the final 161 participants, 78 (48.4%) are women, 81 (50.3%) are men and 2 (1.2%) preferred not to identify themselves or left the answer blank. The mean age was 30.01, with *sd* = 12.5. The sample consisted mainly of workers (40.4%) and students (31.1%), and 23% choose an option "Both a student and an employee". The mean number of years of work experience was 7,2.

Research Design and Materials

In order to test my main hypothesis and see whether exposing people to the threat of social harm will make them perceive more harm and moral disapproval, we developed an online vignette-based experiment. After basic demographic questions, such as age and work experience, participants were randomly assigned either to the experimental or control condition. In experimental condition the participants were exposed to the threat of social harm, in the form of fictional social safety campaign, to maximise ecological validity. Participants in the control condition were not exposed to the posters or any other stimuli. The questionnaire was otherwise identical. To test the effect on two different situations and increase generalizability, we then asked the participants about their perceptions of two unrelated vignettes. We also measured individual differences. While many variables were measured, for the purpose of my research question I only consider harm perception, moral disapproval and internalised moral identity.

Manipulation of Exposure to the Threat Social Harm

To manipulate the exposure to threat of social harm, we created a fictional social safety campaign (Figure 1), presented as four instagram-post-like posters. Social safety campaign was chosen to increase ecological validity, as they are a common way to highlight

the threat and harm social situations can cause. Instagram post was chosen as a realistic medium often used by firms. Overall, all materials are designed in a way that would minimise their novelty, instead focusing on common formats and topics. Both the layout and the content of the fictitious campaign are inspired by the "Just Ask" poster campaign launched by the University of Groningen in April 2023 (University of Groningen, 2023). The aim of the campaign is to make people aware of the invisible harm which can happen in social situations and the threat it can present.

To make it clear to the recipient which ambiguous forms of harmful behaviour the campaign is targeting, the two posters pointing out the potential harm contain speech bubbles with examples of interactions that can be hurtful even without malicious intent. The key message here is that harm can result from verbal interactions and that the assessment of this harm is in the eye of the beholder and does not depend on intentions. The other two posters show standards of behaviour and direct calls to action. They point out the individual's responsibility to recognise and address inappropriate behaviour, which increases the personal relevance of the manipulation, makes it more personal.

Within the questionnaire, to ensure proper attention to the posters, participants were instructed to imagine that the Instagram post was created as part of a social safety campaign implemented by a big firm. They were also asked to consider the goals of the campaign, and briefly summarize the main message of the campaign.

Figure 1

Social Safety Campaign

Vignettes

To estimate the effect of the manipulation on interpretation of ambiguous social situations, we created two fictional vignettes, presented in Figure 2. Both vignettes included actions that could be interpreted as harmful. The vignettes were designed as WhatsApp messages to increase the similarity to real-life digital interactions. WhatsApp is a very popular message exchange platform, and it is very likely that the participants are familiar with it. Moreover, in text messages participants are not able to read body language and facial expressions. They are required to make assumptions about the situation from text alone, making it easier to judge potential effects of experimental condition. Both vignettes are based on situations that should be familiar to most participants. First vignette was about social exclusion in the workplace, henceforth referred to as the exclusion vignette, and second

vignette was about (in)appropriate comments in the workplace, henceforth referred to as the outfit vignette.

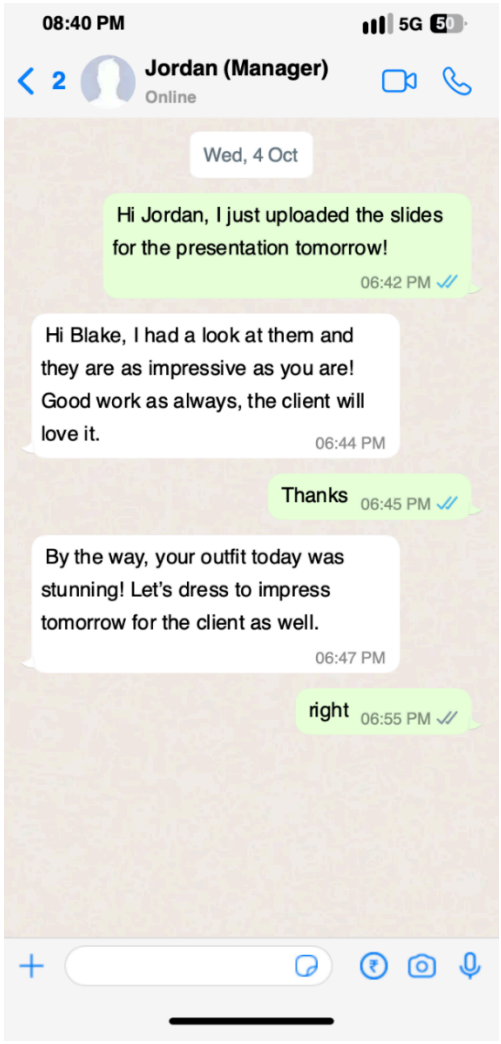
To promote ambiguity, the vignettes were designed in a way that leaves room for interpretation. For instance, in the outfit vignette, the simple reply ‘right’ was added as a last statement to create an open-ended conclusion. Participants must decide themselves whether it was meant as simple agreement or passive aggressivity, signalling being upset. The vignettes have been designed in a way that the harm is not overly explicit. For example, in the exclusion vignette, the exclusion is communicated casually, which could be perceived as either innocent or as deliberate exclusion. Lastly, gender-neutral names were included in both vignettes to lessen the effect of gender bias on the participants’ responses, and to simplify the study’s design. Participants were then asked to answer questions about their interpretation and reactions to the behaviours in the vignette.

Figure 2

Vignettes

Exclusion Vignette

Outfit Vignette



Measures

Perceived Harm

To measure perceived harm, respondents were asked how much harm they thought the person in each vignette experienced, by rating it on a 7-point Likert scale ranging from 0 '*no harm at all*' to 6 '*a great deal of harm'*. This was adapted from Dakin et al. (2023) by our supervisor.

Moral Disapproval

As direct question on moral disapproval might not be a very reliable measure by itself, I am using adapted 3-item scale for moral outrage derived from Skitka, Bauman, & Mullen (2004). This scale asks explicitly about perceived moral wrongness, but also about emotions

that are directly associated with moral disapproval (Bruno et al., 2023), increasing the scope compared to a single explicit question. Respondents were asked to reflect on the team's or person's behaviour displayed in the text messages, by indicating the extent to which they agreed with three statements on a 5-point Likert scale ranging from 1 '*strongly disagree*' to 5 '*strongly agree*'. These statements included '*X's actions made me angry*', '*X's actions are morally wrong*', and '*X's actions upset me*'. In the present study, the scale was highly reliable for both vignettes, with α=0.865 and α=0.945 respectively.

Internalised Moral Identity

Next, I measured internalised moral identity of the participants. I used internalisation sub-scale of moral identity measure developed by Aquino and Reed (2002). The participants were asked to imagine how a person that possessed certain characteristics would think, feel and act. The characteristics were: caring, compassionate, fair, friendly, generous, helpful, hardworking, honest, and kind. Then they were asked 5 questions, such as *"I would be ashamed to be a person who had these characteristics"* or *"Being someone who has these characteristics is an important part of who I am"*. They were asked to rate how much they agree or disagree on a 7-point Likert scale, with 1 = *Strongly disagree* and 7 = *Strongly agree*. Two items were reverse-coded. Higher overall score means higher internalisation of moral identity. The procedure, including the number of answer options and such, is identical to the original measure and its following applications in published studies (Aquino et al., 2009). In the past the scale has shown to be reliable and its Cronbach's alpha ranged from .70 to 0.83 (Lutz et al., 2022). In the present study, the Cronbach's alpha was .741.

Results

Descriptive Statistics

In Table 1, I report descriptive statistics of variables of interest, per vignette per outcome variable per condition.

Table 1

Descriptive Statistics for Harm Perception and Moral Disapproval per Vignette

			Experimental Condition		Control	
Vignette	Variable	Total Mean	Mean	SD	Mean	SD
Exclusion						
	Harm Perception	5.140	5.170	1.312	5.120	1.443
	Moral Disapproval	4.031	4.104	.864	3.964	.902
Outfit						
	Harm Perception	3.140	3.350	1.730	2.950	1.707
	Moral Disapproval	2.708	2.913	1.369	2.520	1.265
Moderator	Moral Identity	6.069	6.126	.679	6.017	.750
	N	161	77	84		

Hypothesis Testing

I proposed two hypotheses as follows. (1) Participants exposed to the threat of social harm will evaluate ambiguous social harms as more harmful and they will show greater moral disapproval, compared to unexposed control group. (2) High internalised moral identity will strengthen the predicted positive relationship between exposure to the threat of social harm, moral disapproval and perceived harm. Each hypothesis is analysed separately, and results are reported per vignette.

Hypothesis 1 Testing: Impact of Threat of Social Harm Exposure on Harm Perceptions and Moral Disapproval

To look at the effects of exposure to threat of social harm, as defined in the first hypothesis, I analysed the mean difference of the two conditions with an independent samples t-test of harm perception and moral disapproval for the two vignettes. In the t-test I assume equal variances on the basis of Levene's test for variance (significance p = .867; .926; .547; .205). T-test statistics can be found in Table 2, means and SD per group can be found in Table 1.

Table 2

Independent Samples T-test of Group Means

		t	df	One-Sided *p*	Mean Difference	Std. Error Difference
Exclusion						
	Harm Perception	-0,228	159	0,410	-0,050	0,218
	Moral Disapproval	-1,001	159	0,159	-0,140	0,140
Outfit						
	Harm Perception	-1,469	159	0,072	-0,398	0,271
	Moral Disapproval	-1,896	159	0,030	-0,394	0,208

Exclusion Vignette. As can be seen in Table 2, for the exclusion vignette, difference of harm perception and moral disapproval between conditions were insignificant. Being exposed to the threat of social harm, in the form of social safety campaign, was not associated with higher scores on harm perception or moral disapproval, not supporting the first hypothesis.

Outfit Vignette. For the outfit vignette, the condition also did not have a significant effect on harm perception. However, exposure to the threat of social harm did have a significant effect on moral disapproval. The analysis showed that compared to those in control condition, participants who were exposed to the threat of social harm morally disapproved of the ambiguous behaviour significantly more at $p = .03$, with effect size of $d = .299$. This partially supports my first hypothesis.

Hypothesis 2: Examining the Moderating Role of Internalised Moral Identity

To test if the main effect is moderated by internalised moral identity, I conducted moderated regression in SPSS with the Process extension. I looked if internalised moral identity would strengthen the predicted positive relationship between exposure to the threat of social harm, moral disapproval and perceived harm. The summary of all models is in Table 3. The models' interactions are reported in Tables 4 to 7.

Exclusion Vignette. The interaction term between the condition and internalised moral identity for harm perception was non-significant, and so was the interaction term for moral disapproval, as can be seen in Table 4 and Table 5 respectively. The R^2 values visible in Table 3, which show the amount of variance of the outcome variable explained by moral identity and experimental manipulation, were also not significant for either outcome variable. These results suggest that the effect of exposure to the threat of social harm on harm perception and moral disapproval is not dependent on the level of internalised moral identity. For this vignette, internalised moral identity was not connected to increase in perceptions of harm or moral disapproval for the experimental manipulation, not supporting hypothesis 2.

Outfit Vignette. Identically to the exclusion vignette, the interaction terms were not significant for either harm perception or moral disapproval, as can be seen in Table 6 and Table 7. Therefore, the overall model fit was also not significant. For this vignette, moral identity was not connected to increase in perceptions of harm or moral disapproval for the experimental manipulation, not supporting hypothesis 2. This indicates that the effect of exposure to the threat of harm on harm perception and moral disapproval is not influenced by levels of internalised moral identity. It is possible that these results were affected by other factors, such as study limitations. This is revisited in the discussion section.

Table 3

Model Summaries: Moderated Regression of Internalised Moral Identity on Condition x Outcome per Vignette

Vignette	Outcome	R	R-sq	MSE	F	df 1	df 2	*p*
Exclusion								
	Harm Perception	.188	.035	1.878	1.909	3	156	.130
	Moral Disapproval	.145	.021	.785	1.122	3	156	.342
Outfit								
	Harm Perception	.129	.017	2.999	.872	3	156	.455
	Moral Disapproval	.193	.037	1.736	2.017	3	156	.114

Table 4

Moderated Regression: Exclusion Vignette: Perceived Harm: Condition x Moral Identity

	Coefficient	Se	t	p	LLCI	ULCI
Constant	4,523	1,215	3,722	<0,001	2,122	6,923
Condition	-2,678	1,882	-1,423	0,157	-6,395	1,039
Moral Identity	0,099	0,200	0,494	0,621	-0,297	0,495
Interaction	0,444	0,308	1,444	0,151	-0,164	1,051

Table 5

Moderated Regression: Exclusion Vignette: Moral Disapproval: Condition x Moral Identity

	Coefficient	Se	t	p	LLCI	ULCI
Constant	3,658	0,786	4,656	0,000	2,106	5,209
Condition	-0,922	1,217	-0,758	0,450	-3,325	1,481
Moral Identity	0,051	0,130	0,394	0,695	-0,205	0,307
Interaction	0,173	0,199	0,868	0,387	-0,220	0,565

Table 6

Moderated Regression: Outfit Vignette: Perceived Harm: Condition x Moral Identity

	Coefficient	Se	t	p	LLCI	ULCI
Constant	1,940	1,536	1,263	0,209	-1,094	4,973
Condition	1,134	2,378	0,477	0,634	-3,564	5,831
Moral Identity	0,168	0,253	0,665	0,507	-0,332	0,669
Interaction	-0,122	0,389	-0,315	0,753	-0,890	0,645

Table 7

Moderated Regression: Outfit Vignette: Moral Disapproval: Condition x Moral Identity

	Coefficient	se	t	p	LLCI	ULCI
Constant	0,837	1,169	0,716	0,475	-1,471	3,145
Condition	2,765	1,810	1,528	0,129	-0,809	6,340
Moral Identity	0,280	0,193	1,451	0,149	-0,101	0,660
Interaction	-0,391	0,296	-1,322	0,188	-0,975	0,193

exposure to the threat of social harm increases the perceived harm and experienced moral disapproval in social interactions. I exposed the participants either to social safety campaign, in the experimental condition, or to no stimuli, in the control condition. Then, I asked participants to evaluate two unrelated ambiguous WhatsApp-style vignettes and asked them about their perceptions of harm and experienced moral disapproval for each vignette.

I hypothesized that exposing the participants to the threat of social harm would lead to higher perceptions of harm and moral disapproval for both vignettes compared to the control condition. This hypothesis was not well supported. For the exclusion vignette, the experimental condition did not lead to statistically significant change in either harm perception or moral disapproval. For the outfit vignette, the experimental condition did not lead to statistically significant change in harm perception. However, participants in the social harm exposure condition did show statistically significant increase in moral disapproval of the outfit vignette, compared to the participants in the control group, with effect size $d = .299$. This is a small, but reasonably high to assume real world implications. Essentially, the results for my first hypothesis are not conclusive, which will be addressed in depth further on.

I also hypothesised that this relationship would be moderated by internalised moral identity, which would be associated with harsher moral disapproval and more severe harm perceptions following the experimental condition. This was not supported for any vignette and for neither outcome variable. In summary, being exposed to the threat of social harm predicted increase only in moral disapproval after the outfit vignette, and moral identity did not moderate the relationship.

It is surprising that for the outfit vignette, moral disapproval is significant while harm perception is not, as I predicted that changes in harm perception and moral disapproval will go hand in hand for each vignette. This assumption was based on previous research (Haidt & Graham, 2007; Schein & Gray, 2018; Turiel, 1983), which was also supported by high

statistically significant correlations between harm perception and moral disapproval within vignettes in the current study. As harm perception for outfit vignette was close to significance ($p = 0.07$) and correlated highly with moral disapproval for said vignette (if controlling for condition, $r = .814$; $p < .001$), it is possible that all that was needed for significance was adequate sample size. However, that is a speculation, and the insignificance also brings forth questions about TDM's claim about perceived harm being a sole foundation for morality evaluation (Schein & Gray, 2018), which is discussed in theoretical implications.

On the other hand, for the exclusion vignette, both outcomes were far from significant, so the results are inconclusive. I predicted that the results would replicate across vignettes. The gap between the results of the vignettes might have been influenced by the specific design of the experimental manipulation. As can be seen in Figure 1, the campaign is focusing on inappropriate comments and the intent behind them, which makes it much more tailored to the outfit vignette, compared to the exclusion vignette. This suggests that exposure to the threat of social harm, for example in campaigns, might be positively associated with increased moral disapproval, but only under certain conditions. One might be, that the evaluated situation must be topically similar to the social harm one was exposed to previously, e.g. it is possible that if the danger of ostracization was highlighted in the campaign, the exclusion vignette would also yield significant results. This is however a speculation, which will be elaborated on later.

It could also be that my hypothesis simply does not hold for situations focused on social exclusion, or that they are generally not viewed as harmful, and I cannot rule out either explanation without further research. For example, when we contrast the topics of the two vignettes, the outfit vignette deals with a topic that is heavily discussed in public discourse. Overall, people agree that there are certain lines that should not be crossed. On the other hand,

social exclusion is not as discussed, and there is a smaller consensus on to what extent are groups obligated to include everyone.

Theoretical Implications

For clarity, I want to highlight that this section is based on the results of the outfit vignette only, as unlike social exclusion, inappropriate comments were mentioned in the campaign. The fact that the campaign did not address topics similar to the exclusion vignette was accidental. It creates a massive limitation for any theoretical interpretation of said vignette, as it means that the experimental manipulation was not done in a way that I intended.

While the results are overall inconclusive, several theoretical implications can be drawn. First, the moderate to high correlations of harm perception and moral disapproval within vignettes further fortifies the support of perceived harm predicting moral judgements, as claimed by mentioned literatures, specifically MFT (Haidt & Graham, 2007) and TDM (Schein & Gray, 2018).

However, when looking at the outfit vignette, which has non-significant difference of perceived harm and significant difference in moral disapproval, it raises a question about the TDM's claim that perception of harm is the sole foundation of moral judgements (Schein & Gray, 2018). It brings forth the idea that perhaps MFT is more accurate it its description of moral foundations. While TDM claims that harm is the sole foundation of moral judgement on which it all depends, MFT says that there are several things that can trigger moral judgement, and that it is possible for us to evaluate something as immoral even if we do not find it harmful (Haidt & Graham, 2007). They claim existence of five moral foundations, and except for harm, they include also purity, fairness, loyalty and respect (Haidt & Graham, 2007). It could be that the experimental manipulation accidentally engaged one of the other

foundations, and harm was not the only reason for the increase in moral disapproval. This could provide further support for MFT pluralistic view of moral bases.

Implications of the Insignificant Moral Identity Moderation

The fact that moral identity did not serve a moderating function, not even on the otherwise significant relationship of the outfit vignette, is surprising. While its non-existent relationship with perceived harm could be explained by harm being a different, albeit to moral disapproval related, construct, the fact that it is not even correlated with moral judgement (viz. Table 8) is bewildering. However, it seems unlikely that all the research predicting the relationship between moral identity and moral disapproval was wrong. Bellow I turn my focus on the measure used, and how social desirability bias might interfere with accurately measuring internalised moral identity.

Internalised moral identity, by definition, is how important being moral is important to one's self-concept. The paper that establishes this subtype of moral identity and defines it, is the same one that developed the internalised moral identity measure used in this study (Aquino & Reed, 2002). Internalised moral identity, measured by this exact measure, was connected to moral behavioural choices (Aquino et al., 2009). While judgements often do not lead to actual behaviour, it is bewildering that internalised moral identity, which is about self-relevance, would lead to moral behaviour but not to corresponding internal evaluations, such as moral judgement in this study. The fact that internalised moral identity was not correlated with moral disapproval raises questions about whether the measure was good choice for the context of this study, or if there could be problem with the measure itself.

The incongruency from the previous paragraph reveals a possible deficiency of the internalised moral identity measure by Aquino and Reed (2002) to fully capture how much being moral is important to one's self concept. Meta-analysis focused on moral identity and its predictive value for moral behaviour found that studies using self-report measures of moral

identity and moral behaviour had higher effect sizes (r =.25) than studies which used implicit moral identity measures and objective or observable criteria for moral behaviour (r =.11) (Hertz & Krettenauer, 2016). These results suggest that people inflate their internalised moral identity due to social desirability bias, which might have happened in this study also. Social desirability might lead to participants inflating their score on internalised moral identity, as they want to feel moral.

But why would internalised moral identity predict behaviour but not judgement? Due to the social desirability bias, internalised moral identity might not actually measure the self-relevance of moral identity accurately. Instead it might accidentally measure something similar to symbolised moral identity, which aims to measure the need for self-presentation as a moral being (Aquino & Reed, 2002). Intuitively people want to be seen as moral, however that may not change their private evaluation of situation, especially if it was ambiguous in the first place. When the measure was developed in 2002, all questionnaires or experiments were most likely filled in person, and not online. In both situations people would want to report high moral identity to feel good about themselves, but in online situation evaluation there is no need to uphold the reported moral identity, as no one can see it. So, it is possible that the measure would, in this context, capture more of a need for moral self-presentation, rather than truly capturing the importance of moral identity to one's self concept.

In conclusion, moral identity might be subject to such a high desirability bias, that the self-importance of the concept might not be possible to capture by the intended measure by Aquino and Reed (2002). It is also possible that since internalised moral identity is naturally skewed to the higher end of the spectrum, the sample size, or the Likert's scale used, were simply not sufficient to create a sufficient spread and properly distinguish between high scorers and low scorers.

Practical Implications

Social harm is lately often mentioned topic, with many organisations investing in enlightenment programs, such as campaigns or workshops. The results of this study suggest that being exposed to the threat of social harm, for example in social safety campaigns, workshops, lectures, etc., can impact the interpretation of ambiguous situations, specifically increase moral disapproval. However, as not all results were significant, it is possible that it happens only under certain circumstances, such when the situation topically aligns with the addressed harm.

While increase in moral disapproval can be beneficial, as moral disapproval can help curb unwanted behaviour by shunning improperly behaving members (Ellemers et al., 2019), it could also have long term negative implications (Kovacheff et al., 2018). While our results are ultimately inconclusive, they give some indication that on individual level, social safety enlightenment programs could increase moral judgement. And this, in turn, could lead to disproportionate organizational and social punishment, which may lead to feelings of injustice, fuelling need for retribution, creating polarization and segregation. (Clifford, 2019; Kovacheff et al., 2018; Skitka et al., 2005; Skitka & Mullen, 2002). Nonetheless, this does not mean that social harm enlightenment programs should be stopped. It is advised to invest in more research on the subject to fully explore the relationship, as the present study has numerous limitations and inconclusive results.

Strengths & Limitations & Future Directions

Strengths

One of the strengths is that the sample is diverse, with the working population and students both, and is spread out age wise. Another strength is that the design, particularly the materials provided the study with high ecological validity, as online interactions and social safety campaigns are commonplace. And while the results of the two vignettes did not align, it

showed valuable data and raised interesting questions for future research, that would otherwise remain hidden.

Limitations & Future Directions

As mentioned in previous sections, this study has numerous limitations. First, the content of the campaign, which served as the experimental manipulation, was focused on inappropriate comments much more than on social exclusion. This made it unintentionally aligned with the topic of the Outfit vignette, as both addressed inappropriate or harsh comments. However, the campaign did not mention the topic of the Exclusion vignette, namely ostracization and social exclusion.

Second major limitation is that the evaluation of harm and moral disapproval of the Exclusion vignette did not predict well the participant's evaluation of the Outfit vignette, even when the participants condition was controlled for (viz exploratory analysis). This points towards questionable design of the vignettes themselves and stresses the need to pretest materials in future research. It could be that the first vignette was simply too ambiguous compared to the second one, as it also depended on the participant's attention, for example on their ability to notice details such as the time frame between the messages.

The study could benefit from a conceptual replication, that would pretest all the materials in advance. To clear up the questions whether the alignment of content of the campaign and the situation matters, it would be interesting to create similar study, but with a 3*2 design, with pretested three social safety campaigns and two pretested vignettes. One campaign would address two topics at the same time, for example inappropriate comments and social exclusion. Then, two campaigns would focus on inappropriate comments and social exclusion respectively. Vignettes would then present ambiguous situation related either to inappropriate comments, or to social exclusion. This would allow us to see if social harm exposure spills over from one type of social harm onto the next, or if the social harm

addressed in the exposure has to be the same as social harm in the ambiguous situation to effect evaluation. It would also show if one of the two topics simply is not affected by exposure to the threat of social harm. At the same time, this design would address the original hypothesis of investigating if exposure to the threat of social harm affects perceptions of harm and moral disapproval.

Third limitation of our study is relatively small sample size, and the fact that for many of the participants, English was most likely not primary language. This opens up the possibility to misunderstand questions, or not fully understanding nuances. In the future, it would be good to recruit at least doubled sample size and administer the questionnaire in participants' native language.

The fourth limitation concerns the measurements used. For harm perception, it was a single explicit question. Having a multiple-item scale in further research could ensure better capture of the concept. Similar can be said for moral disagreement, which was measured by adapted scale of moral outrage, and could benefit by creating a scale unique to moral disagreement. As talked about in general discussion, internalised moral identity measure might be so heavily affected by social desirability, that it might ultimately not measure what it aims to measure. I would recommend detailed revision of the concept and corresponding measure, and creating research design investigating the effect of social desirability on reported internalised moral identity.

Finally, I only measured perceived harm. If other proposed moral foundations were measured in this study, they could have been significant either on their own, or used to improve the model. In future research, I advise to measure more moral foundations other than just harm. There are also possible other research directions about how highlighting harm or danger effects evaluation. For example, the impacts of the current fear-mongering climate on

social media around food, or the effects of negative political campaigns, for example on migration topics.

Conclusion

The polarization in our society is rising and creates significant challenges for cooperation and social cohesion (Kubin & Von Sikorski, 2021). One aspect that seems to be especially pronounced within the polarization are differences in moral evaluations, guided by differences in perceptions of harm (Gray & Kubin, 2024). In this bachelor thesis I set to explore if one of the factors behind these differences in evaluations is exposure to stimuli that highlights the threat of social harm, and if this relationship is strengthened by internalised moral identity. While this research did not provide conclusive results, the results suggest that under certain circumstances, exposure to threat of social harm can be connected to higher moral disapproval, and harm perception is strongly correlated with increasing moral disapproval. Due to numerous limitations, this study should be considered with caution, and in the future should be conceptually replicated in a way that eliminates at least part of the current limitations.

References

Andersen, K., Djerf-Pierre, M., & Shehata, A. (2024). The Scary World Syndrome: News Orientations, Negativity Bias, and the Cultivation of Anxiety. *Mass Communication and Society*, 1–23. https://doi.org/10.1080/15205436.2023.2297829

Aquino, K., Freeman, D., Reed, A., Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, *97*(1), 123–141. https://doi.org/10.1037/a0015406

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Barclay, L. J., Whiteside, D. B., & Aquino, K. (2014). To Avenge or Not to Avenge? Exploring the Interactive Effects of Moral Identity and the Negative Reciprocity Norm. *Journal of Business Ethics*, *121*(1), 15–28. https://doi.org/10.1007/s10551-013-1674-6

Baumert, A., & Schmitt, M. (2009). Justice-sensitive interpretations of ambiguous situations. *Australian Journal of Psychology*, *61*(1), 6–12. https://doi.org/10.1080/00049530802607597

Bellet, B. W., Jones, P. J., & McNally, R. J. (2018). Trigger warning: Empirical evidence ahead. *Journal of Behavior Therapy and Experimental Psychiatry*, *61*, 134–141. https://doi.org/10.1016/j.jbtep.2018.07.002

Bleske-Rechek, A., Deaner, R. O., Paulich, K. N., Axelrod, M., Badenhorst, S., Nguyen, K., Seyoum, E., & Lay, P. S. (2023). In the eye of the beholder: Situational and dispositional predictors of perceiving harm in others' words. *Personality and Individual Differences*, *200*, 111902. https://doi.org/10.1016/j.paid.2022.111902

Bruno, G., Spoto, A., Lotto, L., Cellini, N., Cutini, S., & Sarlo, M. (2023). Framing self-sacrifice in the investigation of moral judgment and moral emotions in human and autonomous driving dilemmas. *Motivation and Emotion*, *47*(5), 781–794. https://doi.org/10.1007/s11031-023-10024-3

Clifford, S. (2019). How Emotional Frames Moralize and Polarize Political Attitudes. *Political Psychology*, *40*(1), 75–91. https://doi.org/10.1111/pops.12507

Dakin, B. C., McGrath, M. J., Rhee, J. J., & Haslam, N. (2023). Broadened Concepts of Harm Appear Less Serious. *Social Psychological and Personality Science*, *14*(1), 72–83. https://doi.org/10.1177/19485506221076692

Ellemers, N., Van Der Toorn, J., Paunov, Y., & Van Leeuwen, T. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366. https://doi.org/10.1177/1088868318811759

Gray, K., & Kubin, E. (2024). Victimhood: The most powerful force in morality and politics. In *Advances in Experimental Social Psychology* (Vol. 70, pp. 137–220). Elsevier. https://doi.org/10.1016/bs.aesp.2024.03.004

Gray, K., Young, L., & Waytz, A. (2012). Mind Perception Is the Essence of Morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012.651387

Haidt, J. (2001). *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*.

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., & Joseph, C. (2008). 19 The Moral Mind: *How Five Sets of Innate Intuitions Guide the Development of Many Culture‐Specific Virtues, and Perhaps Even Modules*. In P. Carruthers & S. Laurence (Eds.), *The Innate Mind, Volume 3* (1st ed., pp. 367–392). Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195332834.003.0019

Haslam, N. (2016). Psychology's Expanding Notions of Harm and Their Moral Basis. In *The social psychology of morality* (pp. 196–214). Routledge. https://doi.org/10.4324/9781315644189.

Haslam, N., Vylomova, E., Zyphur, M., & Kashima, Y. (2021). The cultural dynamics of concept creep. *American Psychologist*, *76*(6), 1013–1026. https://doi.org/10.1037/amp0000847

Hertz, S. G., & Krettenauer, T. (2016). Does Moral Identity Effectively Predict Moral Behavior?: A Meta-Analysis. *Review of General Psychology*, *20*(2), 129–140. https://doi.org/10.1037/gpr0000062

Hester, N., Payne, B. K., & Gray, K. (2020). Promiscuous condemnation: People assume ambiguous actions are immoral. *Journal of Experimental Social Psychology*, *86*, 103910. https://doi.org/10.1016/j.jesp.2019.103910

Hjeij, M., & Vilks, A. (2023). A brief history of heuristics: How did research on heuristics evolve? *Humanities and Social Sciences Communications*, *10*(1), 64. https://doi.org/10.1057/s41599-023-01542-z

Huebner, B., Lee, J., & Hauser, M. (2010). The Moral-Conventional Distinction in Mature Moral Competence. *Journal of Cognition and Culture*, *10*(1–2), 1–26. https://doi.org/10.1163/156853710X497149

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative Information Weighs More Heavily on the Brain: The Negativity Bias in Evaluative Categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887–900.

Kovacheff, C., Schwartz, S., Inbar, Y., & Feinberg, M. (2018). The Problem with Morality: Impeding Progress and Increasing Divides. *Social Issues and Policy Review*, *12*(1), 218–257. https://doi.org/10.1111/sipr.12045

Kubin, E., & Von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, *45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Leavitt, K., Reynolds, S. J., Barnes, C. M., Schilpzand, P., & Hannah, S. T. (2012). Different Hats, Different Obligations: Plural Occupational Identities and Situated Moral Judgments. *Academy of Management Journal*, *55*(6), 1316–1333. https://doi.org/10.5465/amj.2010.1023

Lin, X., & Loi, R. (2021). Punishing the Perpetrator of Incivility: The Differential Roles of Moral Identity and Moral Thinking Orientation. *Journal of Management*, *47*(4), 898–929. https://doi.org/10.1177/0149206319870236

Lutz, P. K., O'Connor, B. P., & Folk, D. (2022). Dimensionality, Item Response Theory, Effect Size Attenuation, and Test Bias Analyses of the Self-Importance of Moral Identity Scale (SIMIS). *Journal of Personality Assessment*, *104*(5), 586–598. https://doi.org/10.1080/00223891.2021.1991359

Midgette, A. J., & Mulvey, K. L. (2024). White American students' recognition of racial microaggressions in higher education. *Journal of Diversity in Higher Education*, *17*(1), 54–67. https://doi.org/10.1037/dhe0000391

Mikhael, S., Watson, P., Anderson, B. A., & Le Pelley, M. E. (2021). You do it to yourself: Attentional capture by threat-signaling stimuli persists even when entirely counterproductive. *Emotion*, *21*(8), 1691–1698. https://doi.org/10.1037/emo0001003

Neta, M., Cantelon, J., Haga, Z., Mahoney, C. R., Taylor, H. A., & Davis, F. C. (2017). The impact of uncertain threat on affective bias: Individual differences in response to ambiguity. *Emotion*, *17*(8), 1137–1143. https://doi.org/10.1037/emo0000349

Osifo, E. (2021, November 4). Support For Anna Faris Is At An All-Time High After Chris Pratt's Instagram Post Thanking His New Wife For His 'Healthy Daughter'. *BuzzFeed News*. https://www.buzzfeed.com/ehisosifo1/support-anna-faris-jack-chris-pratt-instagram-healthy?bfsource=relatedmanual

Schein, C., & Gray, K. (2016). Moralization and Harmification: The Dyadic Loop Explains How the Innocuous Becomes Harmful and Wrong. *Psychological Inquiry*, *27*(1), 62–65. https://doi.org/10.1080/1047840X.2016.1111121

Schein, C., & Gray, K. (2018). 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

Skarlicki, D. P., & Rupp, D. E. (2010). Dual processing and organizational justice: The role of rational versus experiential processing in third-party reactions to workplace mistreatment. *Journal of Applied Psychology*, *95*(5), 944–952. https://doi.org/10.1037/a0020468

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral Conviction: Another Contributor to Attitude Strength or Something More? *Journal of Personality and Social Psychology*, *88*(6), 895–917. https://doi.org/10.1037/0022-3514.88.6.895

Skitka, L. J., & Mullen, E. (2002). The Dark Side of Moral Conviction. *Analyses of Social Issues and Public Policy*, *2*(1), 35–41. https://doi.org/10.1111/j.1530-2415.2002.00024.x

Smetana, J. G., Jambon, M., & Ball, C. L. (2018). Normative Changes and Individual Differences in Early Moral Judgments: A Constructivist Developmental Perspective. *Human Development*, *61*(4–5), 264–280. https://doi.org/10.1159/000492803

Srull, T. K., & Wyer, R. S. (1979). The Role of Category Accessibility in the Interpretation of Information About Persons: Some Determinants and Implications. *Journal of Personality and Social Psychology*, *37*(10), 1660–1672.

Täuber, S. (2019). Moralization as legitimization for ostracism. In S. Rudert, R. Greifeneder, & K. Williams (Eds.), *Current Directions in Ostracism, Social Exclusion, and Rejection Research* (1st ed., pp. 171–189). Routledge. https://doi.org/10.4324/9781351255912-11

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Walsh, J. P. (2020). Social media and moral panics: Assessing the effects of technological change on societal reaction. *International Journal of Cultural Studies*, *23*(6), 840–859. https://doi.org/10.1177/1367877920912257

Wheeler, M. A., McGrath, M. J., & Haslam, N. (2019). Twentieth century morality: The rise and fall of moral concepts from 1900 to 2007. *PLOS ONE*, *14*(2), e0212267. https://doi.org/10.1371/journal.pone.0212267

Appendix A

Correlations Controlled for Conditions

Within-Vignette Correlations

Exclusion Vignette. For Exclusion Vignette, harm perception showed strong positive correlation with moral disapproval, significant at $p < 0.01$. Moral identity was not significantly correlated with either outcome variable.

Outfit Vignette. For Outfit Vignette, harm perception also showed strong positive correlation with moral disapproval, significant at $p < 0.01$. The correlation was stronger than for the other vignette, by .133. Same as for the other vignette, moral identity was not significantly correlated with either outcome variable. This undermines the proposed moderation hypothesis.

Between-vignettes Correlation

Harm Perception x Harm Perception. The between-vignettes correlation of harm perception is significant, but low. I calculated the proportion of explained variance, $R^2 = 0.05$, which shows that harm perception after one vignette was not at all a good predictor for harm perception of the other vignette.

Moral Disapproval x Moral Disapproval. Similarly, the between-vignettes correlation of moral disapproval is also significant but low, with $R^2 = 0.04$. So within this study moral disapproval of first vignette has basically no real predictive value for moral disapproval expressed after the second vignette.

Harm Perception x Moral Disapproval. When looking at this association crosswise between vignettes, the relationship is not significant. So harm perception of one vignette did not predict moral disapproval for the other vignette. This, and the other low between-vignettes correlations, is surprising and its implications are further discussed in the discussion section.