# Seeing Invisible Harm:

# The Impact of Harm Salience & Empathetic Concern on Moral Judgment

Maike Malena Müller-Kuckelberg

S4785398

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group 43

Supervisor: Dr. Maja Graso

Second evaluator: Roxana Bucur

In collaboration with:

Katarina Vastova, Laura Keijzer, Lisette Abels, Lena Kudelska, & Cindy Oosterhuis

01.07.2024

*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**

This study explores how efforts to raise social harm awareness (e.g. social safety campaigns) influence the perception of psychological harm and subsequent moral judgments, and how empathetic concern (EC) may influence this relationship. Psychological harm, often less explicit than physical harm, requires inferring the thoughts and feelings of victims. As the societal understanding of harm broadens to encompass more subtle, psychological phenomena, organizations increasingly make efforts to raise awareness for social safety, e.g. by launching campaigns under this banner. Social safety campaigns aim at defining and denouncing potentially harmful behaviors, thereby perhaps priming individuals to perceive greater harm in ambiguous situations. We conducted an online vignette-based experiment to explore this claim. We predicted that individuals would rate ambiguous, potentially harmful vignettes as more harmful and morally wrong if being exposed to a social safety campaign prior. Further, we expected empathetic concern (EC) to moderate this effect, since EC plays an important role in seeing "invisible" psychological harm and making moral judgments. Findings were largely not statistically significant; however, significant correlations between harm perception and moral judgment as well as an effect of the social safety campaign on increased moral disapproval of a subtle case of sexual harassment ($p = .03$) were found.

*Keywords:* harm salience, moral judgment, empathetic concern, vignette-based experiment, empathy-focused interventions

# Seeing Invisible Harm:

## The Impact of Harm Salience & Empathetic Concern on Moral Judgment

Have you ever told a white lie, that did not hurt anybody? Or perhaps you have lied to avoid hurting someone's feelings? Would you consider such lies to be immoral? While many agree that lying is fundamentally wrong, our moral judgment often hinges on the extent of harm a lie causes. This principle applies to many moral issues, where arguments frequently center around the harm inflicted on a victim. When harm is obvious, moral judgment is often straightforward. For instance, a person who kicks a dog is clearly acting immorally. However, when harm is less evident, moral judgments become more complex. During the 2016 elections, for example, an incident at a US university sparked heated debate. Unknown individuals sprayed the slogan "Trump 2016" on campus walls. While some students viewed this as an act of violence and demanded severe consequences for the perpetrators, others considered the slogan harmless (Haidt & Haslam, 2016). This is just one example of how differing perceptions of harm drive moral divides. More and more debates arise, for example, over language and terminology, comments about persons' physical appearance, or satirical depictions of religious figures, with arguments centered on the potential for certain groups to feel hurt.

Different people perceive different situations as differently harmful. The more apparent the harm in a situation, the more morally reprehensible it becomes, wherefore differences in harm perception fundamentally shape moral disagreement (Gray & Kubin, 2024). While obvious physical harm is fairly evident, subtle forms of harm are more ambiguous. Harm-related concepts are increasingly understood to not only encompass blatant, physical harm, but also more nuanced forms of psychological harm (Haslam et al., 2020). Psychological harm is often less explicit than physical harm, leaving room for subjective evaluations. It can therefore be

considered “invisible harm”, because psychological wounds arise in the minds of the victims and are difficult for outsiders to grasp. To better understand moral divides, we must explore the factors that cause people to perceive harm differently in such ambiguous cases.

Although many factors contribute to harm perception, our study focuses on two elements. We examine social safety campaigns as a tool to increase the salience of harm in ambiguous situations. Such campaigns are one of many measures that are increasingly implemented by organizations to raise awareness for the potential harmfulness of everyday interactions. Additionally, we investigate the role of empathetic concern (EC), which describes an increased awareness of others' suffering (Wispé, 1991) and, therefore, attunes individuals to detect “invisible” psychological harm. Our study aims to enhance the current literature on harm and morality by investigating how institutional efforts to raise harm awareness, along with individual differences in sensitivity to harm, influence moral judgments.

## Theoretical Foundation

### Harm, Norms & Morality

First, we need to establish why changes in harm salience impact moral judgments. An extensive body of literature argues that an act is seen as immoral if it causes counter-normative harm to a victim (Schein & Gray, 2018). According to Gray & Kubin (2024), an evolutionary perspective provides clues as to why harm and morality are so closely linked. According to them, it is human nature to be afraid of being victimized. In this context, a victim is an innocent agent who is unjustly harmed by an aggressor (Gray & Kubin, 2024). According to Boehm (1982), morality has the evolutionary function to reduce group conflict by encouraging cooperation and sanctioning aggressive, harmful or antisocial behaviors. Those include but are not limited to direct harm (i.e. pain and suffering). Moral Foundations Theory (MFT; Graham et al., 2013)

identifies five different domains of threat that morality has evolved to denounce: direct harm, unfairness, disobedience, disloyalty and impurity. Gray & Kubin (2024) argue that these "moral concerns" universally revolve around the infliction of harm, whether it be upon individuals or the collective welfare of a group. However, harm is deemed immoral only when occurring counter-normative, i.e. when a victim is unjustly harmed without any group norms to justify it (Gray & Kubin, 2024). Norms are, therefore, essential cultural frameworks that help us to make moral judgments and assess the harmfulness of an act.

According to the Theory of Dyadic Morality (TDM; Schein & Gray, 2018), the relationship between harm and morality is bidirectional. Simply put, we perceive immoral acts as harmful, and harmful acts as immoral. This interaction creates a feedback loop, termed the "dyadic loop", where moral disapproval and the perception of harm constantly reinforce each other. From this we derive the assumption for our study, that the more obvious harm is in a situation, the more it is morally condemned. However, the subjective nature of psychological harm makes harm assessments ambiguous.

**Psychological Harm and Social Safety**

Harm extends beyond the physical realm, encompassing psychological and emotional dimensions. According to Nick Haslam (2016), over the past decades, the psychological concept of harm has undergone a gradual semantic expansion, beyond its predominantly physical connotation. Scholars identify two dimensions of harm-concept creep: vertical concept creep refers to the phenomenon of lowering the threshold of what constitutes harm gradually, to encompass less severe phenomena. Horizontal concept creep describes the conceptual broadening of the harm notion to include quantitively more phenomena (Haslam et al., 2020). Thus, verbal, non-physical interactions are nowadays increasingly ascribed the potential to be

psychologically harmful. However, psychological harm is often less explicit, as observers must infer the thoughts and feelings of a victim (Yoo & Smetana, 2019). Hence, the question of what constitutes harm is increasingly a matter of subjective interpretation rather than "externally verifiable, tangible" criteria (Bleske-Rechek et al., 2023, p. 1). While advocates emphasize the potential of expanded harm concepts to make previously overlooked, subtle aggression visible, critical voices express their concerns about an inflation and dilution of the harm notion and the consequent potential for conflict in moralized debates (Haidt & Haslam, 2016).

Regardless of whether this process is viewed positively or negatively, changing the definition of harm has tangible effects on societies. Ian Hacking (2007) claims, that once a concept is defined, it becomes real and is perceived as such. An example of this is the implementation of social safety policies. As non-physical social interactions are increasingly recognized as potentially harmful, institutions react by implementing guidelines addressing corresponding behaviors. A buzzword gaining increasing traction in this context is "social safety," which refers to the idea of creating safe environments by minimizing psychological stressors that potentially harm individuals (Slavich, 2020).

The University of Rotterdam, for example, states on its website: "*Social safety means that we treat each other with respect and do not harass, discriminate or intimidate each other. It means that we do not tolerate transgressive behavior"* (University of Rotterdam, 2024). One way in which organizations communicate the importance of this issue is by campaigning. To quote the University of Amsterdam, the university's most recent social safety campaign aims at *"raising awareness of, and helping to recognize and name undesirable behavior, for and by everyone in the organization"* (University of Amsterdam, 2024). Our University in Groningen has also launched the "Just Ask" campaign that is *"intended to contribute to the continued*

*strengthening of an open and safe culture within the UG"* (University of Groningen, 2023). Social Safety Campaigns are, hence, one of many measures taken by organizations to navigate an increasingly subjective understanding of harm.

**Increasing Harm Salience through Social Safety Campaigns**

There are several studies that allow us to anticipate that social safety campaigns increase the salience of psychological harm in ambiguous scenarios. Specifically, Bleske-Rechek et al. (2023) have demonstrated in a study on verbal microaggressions how brief informational texts about the potential harm elicited by others' words led participants to perceive ambiguous sentences like *"That's different from what you usually wear"* as more harmful. With regard to the US terror alert system, it has also been shown that the increased salience of potential threat results in negative affective responses in anticipation of harm (McDermott & Zimbardo, 2006). Similar claims about increased harm sensitivity have been extensively made about trigger warnings. For example, Filipovic (2014) argued in The Guardian that the growing use of trigger warnings at universities, intended to mentally prepare students before seeing distressing content, might backfire. Instead of safeguarding students, it could make them overly fragile and ill-prepared for the real world, where upsetting events often occur without any warning signs (Filipovic, 2014). However, among the scientific community, such statements remain strongly disputed (Bridgland et al., 2023).

Further, we argue that social safety campaigns potentially increase the harmfulness of acts by framing them as immoral. Following the logic of TDM (Schein & Gray, 2018), counter-normative acts are perceived as immoral and thus harmful. Injunctive social norms can be understood as implicit rules of how socially acceptable certain behaviors are. If these norms do not provide justification for a specific harmful behavior, it is perceived as immoral (Warner et

al., 2022). This explains why a man knocking another unconscious on the street is seen as immoral, but not when it happens in a boxing ring in front of spectators. Evidence suggests that campaigns effectively shift injunctive social norms, tackling harmful behaviors such as tobacco use (Lahiri et al., 2021), unsafe sexual practices (Kesterton & Cabral De Mello, 2010) or risky driving behavior (Liu et al., 2022). While most research in the field is concerned with direct, physical harm, recent pilot studies also address non-physical harm, for example, campaigns tackling injunctive norms regarding women's public and political participation in Burundi, Rwanda, Somalia, and Sudan (Kakal et al., 2020). Based on the aforementioned examples of social safety campaigns (Amsterdam, 2024; Groningen, 2023; Rotterdam, 2024), we argue that by naming and speaking out against undesirable behaviors, social safety campaigns define injunctive social norms within organizations. As a result, counter-normative behavior is framed as immoral and, thus, perceived as more harmful (Schein & Gray, 2018).

The final reason we anticipate that social safety campaigns increase the salience of harm is their agenda-setting function in defining what is considered harmful. By calling upon recipients to identify and name undesirable and harmful behaviors (Amsterdam, 2024; Groningen, 2023; Rotterdam, 2024), we argue that they have the potential to broaden recipients' concepts of harm. As Haslam et al. (2020) claim, concept creep can, in part, be understood as a motivated process. The term "expansion agents" refers to deliberate actors fueling concept inflation to amplify the perceived seriousness of and moral responses to social issues (Haslam et al., 2020). Campaign slogans can act as expansion agents by defining what is to be understood as harmful (e.g. "Words can hurt"), therefore, morally relevant.

**Empathetic Concern as a Moral Force**

In the preceding paragraphs, we have established the close link between harm and moral

disapproval. However, we have also found that psychological harm is not always equally obvious to everyone, but instead depends on subjective interpretation. We consequently argue that a fundamental prerequisite for moral judgment lies in the capacity and motivation to detect others' suffering – in other words to *empathize*. This is especially relevant with regard to psychological harm, as it requires inferences about feelings and thoughts of a victim (Yoo & Smetana, 2019). The inner world of others is inaccessible to us; we can never fully grasp their feelings or the extent of their suffering. Philosophers have coined the term *"The problem of the other mind"* to describe this fundamental human deficit (Gray & Kubin, 2024).

Empathy, conversely, can be "understood as the primary epistemic means for gaining knowledge of other minds" (Stueber, 2019, p. 03); i.e. the psychological mechanism that allows us to draw conclusions about others' subjective experience. Psychologist Edward Titchener coined the term "empathy" as a translation of the German word "einfühlen" ("feeling into"), referring to the experience of congruent emotions of one's counterpart and responding to the others' possible distress (Slote, 2003). Although empathy has been extensively shown to be related to prosocial behavior (Decety & Cowell, 2014), the connection to moral judgment is not as straight forward. Instead, one facet of empathy seems to be involved in judgments about moral transgressions: "sympathy".

In his Moral Philosophy, 18th century Enlightener David Hume operates with the term sympathy to describe a *"heightened awareness of the suffering of another person as something that needs to be alleviated"* (Wispé, 1991, p.68). More than a mere "feeling into", Hume's sympathy manifests itself in the acknowledgement of suffering and the desire to alleviate it (Stueber, 2019). Hume's notion of sympathy may be best reflected in the modern day concept of "Empathetic Concern" (EC), a subscale in Davis' (1983) "Interpersonal Reactivity Index" (IRI).

Empathetic Concern has been found to be related to increased moral judgments about psychological harm (Ball et al., 2017) and to moderate the effect of harm-salience on moral judgments (Yoo & Smetana, 2019). According to previous research, individuals with high EC perceive more harm and make stronger moral judgments, even when harm salience is low (Ball et al., 2017; Yoo & Smetana, 2019). We suspect that, in our study too, individual differences in EC may moderate the effect of increased harm salience on the perception of harm and moral disapproval of ambiguous scenarios. The effect of the social safety campaign should therefore be particularly strong for low EC individuals, as high EC would lead participants to perceive more harm and moral wrongness in ambiguous scenarios also without being exposed to a campaign. We argue that this can be explained by a heightened harm awareness of high EC individuals. Several studies suggest why this might be.

First of all, high EC individuals tend to hold broader concepts of harm (Haslam et al., 2020). Although Haslam et al. (2020) understand concept creep as a universal phenomenon, Bleske-Rechek and colleagues' (2023) study on verbal microaggressions suggests between-subject variation in concept breadth. Haslam et al. (2020) further assume that, universally, concept creep is driven by an increase in harm-based morality in Western cultures, indicating that individuals who strongly endorse the moral foundation of harm (Graham et al., 2013) adopt broader concepts of harm more readily (McGrath et al., 2019). McGrath et al. (2019) further find endorsement of harm-based morality to be strongly associated with high levels of EC. This brings us to the conclusion, that high EC individuals tend to hold more inclusive concepts of what is considered harmful.

Resulting from broader concepts of harm, we secondly argue that high EC individuals are more aware of the potential harmfulness of social interactions, as they classify more behaviors as

being harmful overall. Consequently, the dyadic loop (Schein & Gray, 2018) causes them to morally disapprove of a broader array of interactions. The expanded perception of harm and the moral condemnation of a wider range of behaviors, driven by harm-based morality, then manifests itself in corresponding norms of social interactions upheld by individuals with strong empathetic concern (Quissell, 2022).

### Present Study

In this study, we aim at exploring the effect of harm salience and individual differences in empathetic concern on harm perception and moral disapproval of ambiguous psychological harm. Harm salience is, in this study, manipulated by exposure to a social safety campaign. We anticipate finding the following effects:

**Hypothesis 1:** Exposure to a social safety campaign will increase the perceived harmfulness and moral disapproval of ambiguous, potentially harmful social interactions.

**Hypothesis 2:** The effect of the social safety campaign on harm perception and moral disapproval will be moderated by individual differences in empathetic concern, as perceived harmfulness and moral disapproval will be generally higher in high EC individuals that are not exposed to a social safety campaign.

## Methods

### Participants & Procedure

To test our two hypotheses, we conducted an online vignette-based experiments with one experimental condition and a control group. The idea was to expose subjects to ambiguous situations that potentially entailed psychological harm. We aimed to determine how harmful and morally reprehensible participants find these situations either with or without being exposed to a social safety campaign. The campaign was designed to increase the salience of harm in the ambiguous vignettes. Additionally, we measured individual differences in empathetic concern to

understand how these differences might moderate the effect of harm salience on harm perception and moral judgment.

Before conducting our experiment, we designed the fictional social safety campaign and the ambiguous vignettes (discussed in more detail below). Our study was then approved by the ethics committee of the University of Groningen, and we started to recruit participants. We recruited participants through our personal networks, for example on platforms like LinkedIn and WhatsApp. Additionally, we recruited a sample on Prolific. In total, we recruited 227 participants, of which 51 were excluded in a preparatory cleaning due to unfinished answers or withdrawal of consent. The final sample used in the analysis consisted of 161 participants, after excluding another 15 participants for missing attention checks. Out of these, 78 (48.4%) were women, 81 (50,3%) men and 2 (1.2%) preferred not to identify themselves. Participants' mean age was 30.02 with a standard deviation of 12.50. Participants were randomly assigned to either the control group (n=84) or experimental condition (n=77). The sample consisted mainly of employees (40.4%) and students (31.1%). Another 23% were both, a student and an employee. The survey was administered via an online form on Qualtrics, with answers collected in English.

**Ambiguous Harm Vignettes: WhatsApp Chats**

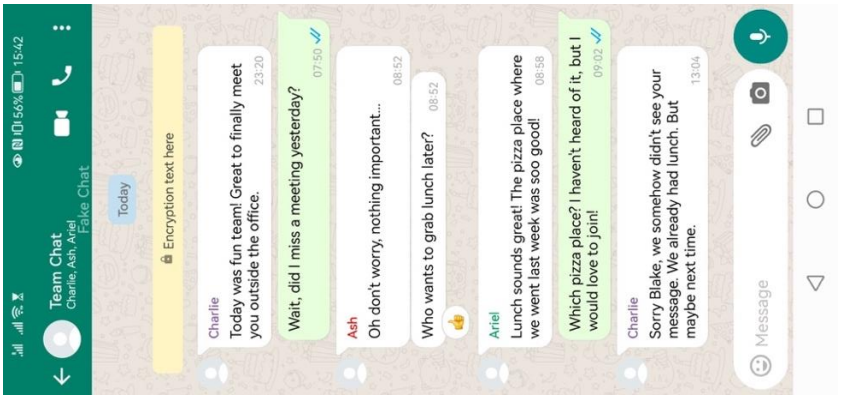
All participants (n = 161) were exposed to two harm-related vignettes (Figure 1). We developed two vignettes detailing potentially harmful, non-physical social interactions at the workplace. The vignettes were designed as WhatsApp chats to increase ecological validity, as we can assume that most participants regularly use some form of instant messaging. WhatsApp is a very popular message exchange platform, and it was very likely that the participants are familiar with it. The chat format allowed participants to immerse themselves in the situation from a first-person perspective, reading the messages as if they were the alleged victim. Also, by using two

different scenarios we intended to cover different types of psychological harm. The first vignette (groupchat vignette) depicted a potential case of bullying, while the second chat (outfit vignette) suggested a form of verbal sexual harassment. Topics such as (in)appropriate compliments from colleagues and social exclusion were intended to touch upon themes that participants are very likely to have experienced either themselves or peers at some point in their lives.

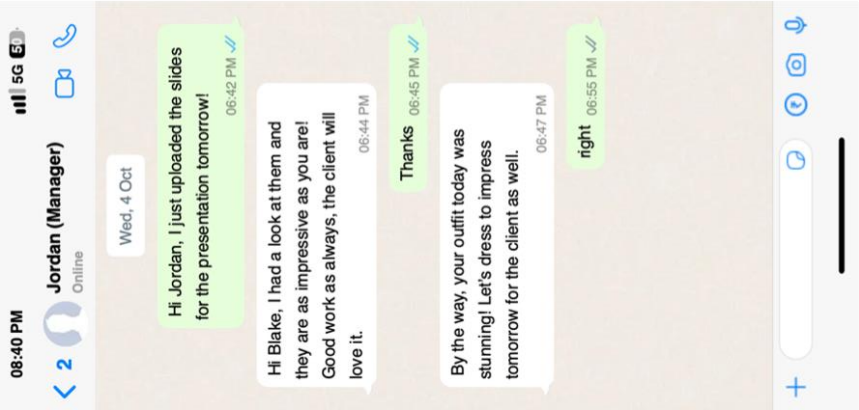
Several components have been incorporated into the vignettes to optimize ambiguity. Firstly, the vignettes were designed in a way that leaves room for interpretation. For instance, no emojis were used that could give hints about the emotions of the alleged perpetrators and victims. Also, because the vignettes were chats, participants were not able to read body language and facial expressions. This way participants are required to rate intent and tone from text alone. Moreover, gender-neutral names were used in both vignettes to lessen the effect of gender bias on the participants' responses and to simplify the study's design. Furthermore, the vignettes have been designed in a way that the harm is not overly explicit. For example, in the vignette that describes a team allegedly excluding or bullying one team member, the exclusion could be perceived as either innocent or as deliberate exclusion. The design of our ambiguous vignettes was inspired by Dakin and colleagues (2023).

**Figure 1**

*Vignette 1: Groupchat*



*Vignette 2: Outfit*




**Harm Salience Manipulation: Social Safety Campaign**

Only the experimental group (n = 77) saw a fictional social safety campaign before reading the chats. This served to increase the harm salience for this group, i.e. to point out the potential harmfulness of social interactions and to emphasize that the fictitious organization is committed to social safety. This fictional social safety campaign consisted of four different posters (Figure 2). A short introductory text invited participants to take a close look at the posters and imagine that the campaign had been launched by their organization or institution. The posters were designed in the typical square format of Instagram posts, as the social media platform is a realistic channel that institutions use to spread social safety campaigns. Both the layout and the content of the fictitious campaign were inspired by the "Just Ask" poster campaign launched by the University of Groningen in April 2023 (University of Groningen, 2023). The slogans stand out in white lettering against a red background, which gives them a warning appearance at first glance. The aim of the campaign is to make people aware of the potential harm that may arise from social interactions ("Words can hurt"; “Didn’t mean it?”) and to define organizational norms of behavior ("Stand firm, speak out"; “Don’t ignore the signals”).

The two posters pointing out the potential harm of social interactions contain speech bubbles with examples of interactions that can be hurtful even without malicious intent. This makes it clear to the recipient which ambiguous forms of harmful behavior the campaign is targeting. 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 bad intentions. The other two posters entail standards of behavior and direct calls to action. They point the individual's responsibility to recognize and address inappropriate behavior out.

**Figure 2**

*Harm Salience Manipulation: Social Safety Campaign*


**Measures**

***Harmfulness***

After reading each vignette, respondents were asked how much harm they thought the alleged victim experienced by rating it on a 7-point Likert scale ranging from 1 *'no harm at all'* to 7 *'a great deal of harm'*. This measure was based on Dakin and colleagues (2023).

***Moral Disapproval***

After each vignette, participants' moral disapproval of the acts was measured. Moral disapproval was measured using a three-item scale for moral outrage derived from Skitka, Bauman, & Mullen (2004). The reasoning behind using a moral outrage scale was to increase internal validity of the measure. Instead of only asking participants about the moral wrongness of

an act, the three-item measure for moral outrage also encompasses emotions related to moral disapproval, such as anger. Respondents were asked to reflect on the team's or person's behavior 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*' An example for one item is *"(Name's) actions make me angry".* We found very good reliability for or moral outrage measure (group chat vignette: α = .87; outfit vignette: α = .95). See Appendix A for a list of all three items.

***Empathetic Concern***

Participants' trait empathetic concern was measured using the 7-item subscale "Empathetic Concern" of the "Interpersonal Reactivity Index (IRI)" (Davis, 1983). Participants' responses to 7 items were measured on a 5-point Likert scale from 1 '*strongly disagree*' to 5 '*strongly agree*'. One item was for example *"I often have tender feelings for people less fortunate than me".* Reliability of the scale was very good (α=0.80). See Appendix B for a list of all 7 items.

***Demographics***

Demographic measures included the gender, age, employment status, work experience in years, and political orientation of participants. Demographical data was collected, but not considered in the analysis.

## Results

In Table 1, overall descriptive statistics (means and standard deviations) and correlations for empathetic concern (EC), perceived harmfulness (Harm) and moral disapproval (Moral) can be found. This shows that the means of harmfulness and moral disapproval in both vignettes are slightly higher in the experimental condition than in the control group. Furthermore, the

significant correlations between the dependent variables are striking. Results are presented for both vignettes (group & outfit) separately.

**Table 1**

*Descriptives and correleations*

| | | Means (standard deviations) | | | | Pearson Correlations | | |
|---|---|---|---|---|---|---|---|---|
| | N | Total | Control | Social Safety | Mean Differ. | EC | Harm group | Moral group |
| EC | 160 | **3.90** (.65) | **3.84** (.74) | **3.97** (.59) | .13 | 1.0 | | |
| Harm group | 161 | **5.14** (1.38) | **5.12** (1.44) | **5.17** (1.31) | .05 | .00 | 1.0 | |
| Moral group | 161 | **4.03** (.88) | **3.96** (.90) | **4.10** (.86) | .14 | .09 | .68* | 1.0 |
| Harm outfit | 161 | **3.14** (1.72) | **2.95** (1.71) | **3.35** (1.73) | .40 | .10 | .22* | .14 |
| Moral outfit | 161 | **2.71** (1.33) | **2.52** (1.26) | **2.91** (1.37) | .39 | .14 | .14 | .21* |

*Note.* *p < 0.01 (2-tailed); Harm group & Harm outfit = harmfulness groupchat vignette & outfit vignette (measured on 7-point Likert scale); Moral = moral disapproval (measured on 5-point Likert scale); EC = empathetic concern measured on 5-point Likert scale

**Hypothesis 1: Impact of Harm Salience on Harm Perception and Moral Disapproval**

First, we investigated whether the group that saw the social safety campaign (n = 77) found the two vignettes more harmful and more morally reprehensible than the control group (n = 84). For that, we assumed equal variances of the group means after consulting Levene's Test for Equality of Variances (Table 2). We then conducted an independent samples t-test to determine whether the group means differed statistically significantly (Table 2). Below, we present the results separately for the group chat vignette and the outfit vignette.

***Groupchat Vignette***

For the groupchat vignette, we found no support for our first hypothesis (Table 2). This means that the participants who were primed with the social safety campaign did not find the behavior depicted in the vignette significantly more harmful or more morally reprehensible.

***Outfit Vignette***

For the outfit vignette, we found no significant difference between the group means for the harmfulness of the behavior. However, the participants who saw the social safety campaign found the depicted behavior significantly more morally reprehensible (Table 2). Therefore, our first hypothesis was partly supported for the outfit vignette.

**Table 2**

*Independent Samples Test*

| | Lavene's Test of Equal Variances | | t-test for Equality of means | | | | |
|---|---|---|---|---|---|---|---|
| | F | sig | Mean difference | Std. Error difference | t | df | p |
| Harm group | .03 | .87 | -.050 | .22 | -.23 | 159 | .41 |
| Moral group | .01 | .93 | -.14 | .14 | -1.00 | 159 | .16 |
| Harm outfit | .37 | .55 | -.40 | .27 | -1.47 | 159 | .07 |
| Moral outfit | 1.62 | .21 | -.40 | .21 | -1.90 | 159 | .03* |

*Note.* N = 161; *p (one-sided) < .05

**Hypothesis 2: Interaction of Experimental Condition with Empathetic Concern**

To explore the predicted interaction effect, we conducted a moderated regression with Process Macro (Hayes, 2013). We analyzed the dependent variables harmfulness and moral disapproval for the groupchat and outfit vignette separately. To do this, we inserted the dependent variables one after the other as Y and the experimental condition (social safety vs. control) as X in the process macro plug-in. We selected empathetic concern (EC) as the

moderating variable W in each case. The program then provided us with 4 interaction models of the form *Y = X x W x Int(X x W)*. We will present the models individually, divided by vignette and dependent variable.

***Groupchat Vignette***

As shown in Table 3, the interaction model does not explain any differences in how harmful participants found the behavior depicted in the group chat vignette. The interaction of the experimental condition with empathetic concern is not significant. Furthermore, Table 4 shows that the interaction model also does not explain the differences in moral disapproval of the behavior in the groupchat vignette. Here too, the interaction of the condition with EC is not significant.

***Outfit Vignette***

Tables 5 and 6 show that the interaction model does not significantly account for different ratings of harmfulness and moral disapproval for the behavior depicted in the outfit vignette. For both dependent variables, the interaction between condition and EC is not significant.

**Table 3**

*Moderated Regression Harm Groupchat Vignette: condition x EC*

| | coefficients | SE | t | p | $R^2$ | F | p |
|---|---|---|---|---|---|---|---|
| condition | .47 | 1.39 | .34 | .73 | | | |
| EC | .04 | .22 | .17 | .86 | | | |
| condition x EC | -.11 | .35 | -.31 | .76 | | | |
| Model Summary | | | | | .00 | .05 | .99 |

*Note.* degrees of freedom: df1 = 3; df2 = 156

**Table 4**

*Moderated Regression Moral Disapproval Groupchat Vignette: condition x EC*

| | coefficients | SE | t | p | $R^2$ | F | p |
|---|---|---|---|---|---|---|---|
| condition | .61 | .89 | .69 | .49 | | | |
| EC | .16 | .14 | 1.17 | .24 | | | |
| condition x EC | -.12 | .22 | -.55 | .58 | | | |
| Model Summary | | | | | .02 | .81 | .49 |

*Note.* degrees of freedom: df1 = 3; df2 = 156

**Table 5**

*Moderated Regression Harm Outfit Vignette: condition x EC*

| | coefficients | SE | t | p | $R^2$ | F | p |
|---|---|---|---|---|---|---|---|
| condition | .35 | 1.72 | .20 | .84 | | | |
| EC | .23 | .27 | .86 | .39 | | | |
| condition x EC | .01 | .43 | .02 | .99 | | | |
| Model Summary | | | | | .02 | 1.12 | .34 |

*Note.* degrees of freedom: df1 = 3; df2 = 156

**Table 6**

*Moderated Regression Moral Disapproval Outfit Vignette: condition x EC*

| | coefficients | SE | t | p | $R^2$ | F | p |
|---|---|---|---|---|---|---|---|
| condition | -.01 | 1.31 | -.01 | .99 | | | |
| EC | .23 | .21 | 1.20 | .27 | | | |
| condition x EC | .10 | .33 | .29 | .77 | | | |
| Model Summary | | | | | .04 | 2.15 | .10 |

*Note.* degrees of freedom: df1 = 3; df2 = 156

Thus, we conclude for both vignettes that the effect of the condition on neither dependent variable was significantly moderated by EC. In other words, how empathetically concerned the participants were did not affect how strongly the campaign influenced their ratings of harmfulness and moral disapproval. This contradicts our hypothesis that the effect of the harm

salience condition would be particularly strong for low EC individuals.

## General Discussion

Thinking back to the incident at the US university that was mentioned in the beginning of this paper, we recall that some students interpreted the slogan “Trump 2016” to be an act of violence, while others judged it harmless (Haidt & Haslam, 2016). Our study aimed at exploring factors that contribute to such differences in harm perception when the amount of harm cannot be objectively quantified but must be subjectively interpreted. For instance, consider the WhatsApp vignettes in our study: determining whether and to what extent a victim is harmed in those virtual conversations is ambiguous. Exploring the factors that are linked to differences in harm perception is relevant to understand moral divides, as we theorized that moral judgments are fundamentally driven by harm. We predicted that social safety campaigns would increase harm perception and moral disapproval by highlighting harm (Bleske-Rechek et al., 2023) and broadening harm concepts (Haslam et al., 2020). Moreover, we suggested an interaction with empathetic concern because EC is related to an endorsement of harm-based morality (Schein & Gray, 2018), broader concepts of harm (Haslam et al., 2020) and has been previously shown to moderate the effect of harm salience on moral judgment (Ball et al., 2017; Yoo & Smetana, 2019).

Although we expected our hypotheses to apply equally to both vignettes, we analyzed them separately. Unlike hypothesized, we found different results for the vignettes. While for the groupchat vignette the groups did not differ in either ratings of harmfulness or moral disapproval of the depicted behaviors, we found a significant difference between group means for moral disapproval in the outfit vignette. However, contrary to our hypothesis, this did not apply to the harmfulness of the outfit vignette. Our theory assumes a close conceptual link between harm and

morality, with TDM even claiming harm to be the fundamental basis of moral judgment (Schein & Gray, 2018). This link was also supported by moderate to strong correlations between the dependent variables in our results. We were, therefore, surprised by the fact that only moral judgment turned out to differ significantly.

Furthermore, we expected empathetic concern to moderate the effect of harm salience on harm perception and moral judgment, as we hypothesize that EC plays a crucial role in recognizing the suffering of others. Based on this, we assumed that high EC individuals feel more harm and moral disapproval even without being explicitly made aware of the harm by a campaign. This hypothesis was not supported by our results, as we did not find a significant interaction of the experimental condition with EC for either of the two vignettes or any of the dependent variables. Despite some surprising and predominantly statistically non-significant results, this study still invites some theoretical implications and recommendation for future research.

**Theoretical Implications**

Firstly, the fact that we found a significant effect only for the outfit vignette contradicted our theory that the social safety campaign would make the harm in both vignettes equally salient. In retrospect, however, we realized that we had unintentionally designed the campaign in such a way that it was particularly well-tailored to the outfit vignette. The outfit vignette suggests a subtle form of sexual harassment and the campaign includes statements as *"Zero tolerance for harassment"* and *"Silence supports harassment"* (Figure 2). Social exclusion as depicted in the groupchat vignette is not explicitly mentioned in the campaign. We consequently conclude that thematic overlap between campaign and vignette is a relevant factor to consider in the design. This means that to effectively highlight the harm in a vignette, the experimental manipulation

should directly relate to the specific form of harm depicted in the vignette. Our hypothesis that a manipulation pointing to harm in general would be sufficient to increase harm salience, therefore, maybe too simplistic. This leads to a significant theoretical implication for future studies that aim to increase harm salience through experimental manipulation (e.g. social safety campaigns): it is crucial to ensure that the manipulation aligns with the specific type of harm being examined.

Secondly, we predicted that both moral disapproval and perceived harmfulness would be significantly higher in the social safety group, based on Schein and Gray's (2018) theory that moral judgment is based on harm perception. Accordingly, a significant effect for moral disapproval should only occur in conjunction with a significant effect for harm. However, we were surprised to find a significant effect only in moral disapproval of the outfit vignette. If our results accurately reflect true effects and are not due to methodological artifacts, they challenge the claim made by TDM that harm is the prerequisite for moral judgment. This, in turn, supports scholars who argue that immorality is not necessarily linked to harm. MFT (Graham et al., 2013), for example, distinguishes five foundations of morality, of which “harm” is only one domain. This means that in this "pluralistic view," moral disapproval can also arise from the violation of any of the other four moral foundations: unfairness, disobedience, disloyalty, or impurity. Although Gray & Kubin, (2024) argue that all five foundations proposed by Moral Foundations Theory (MFT) can be traced back to harm, MFT understands them to be independent foundations. Thus, TDM and MFT present conflicting perspectives. TDM asserts that moral judgment is always linked to harm (Schein & Gray, 2018), with the five moral foundations representing different facets of harm. From this perspective, “disobedience” is harmful to the structure of a group, and since the evolutionary role of morality is to reduce group

conflict, it denounces this form of harm. In contrast, MFT proposes that the moral foundations are independent concepts, and that moral disapproval is not necessarily linked to harm. Therefore, according to MFT, there can be "harmless wrongs"; e.g. consensual incest that is immoral due to impurity but does not directly harm anyone. If, in the case of the outfit vignette, only moral disapproval significantly increased, it would support MFT's claims, suggesting that the cause of moral judgment might be something other than harm. However, we found moderate to high correlations between our dependent variables for both vignettes, which we understand as support for the claim that harmfulness and moral disapproval increase together. Since perceived harm of the outfit vignette is bordering statistical significance ($p=.07$), the results could be as well attributed to lacking power due to sample size.

Thirdly, we argue for a link between the predominantly unsignificant results of our first hypothesis and the considerations of our second hypothesis. We hypothesized that individuals with high empathetic concern would perceive the vignettes as more harmful and morally wrong even without being exposed to the campaign. We accordingly expected the effect of the campaign to be especially pronounced for low EC participants. Our data, however, suggests generally high levels of empathetic concern in our sample, with a mean of 3.9 on a 5-point Likert scale. Only 5 participants scored lower than 3, while 84 individuals had EC scores of 4 or above. Considering our second hypothesis and the predominantly high levels of empathetic concern, we argue that our results are not necessarily contradicting our theoretical foundation. Eventually, significant effects could be found with the design that we used, if only participants had lower EC. Perhaps this underrepresentation of individuals with low EC accounts for the non-significance of the tested interaction. However, it remains counterintuitive that EC seems to be not or only weakly correlated to our dependent variables. This raises the question of whether our

measure for EC reflects how empathetically concerned our participants really are. We discuss the potential for social desirability bias due to self-report measures in more detail in the limitations section.

**Limitations and Future Directions**

The first limitation of our study is that we did not have the capacity to pretest our vignettes and the social safety campaign. We designed them based on face value, inspired by ambiguous harm vignettes used by Dakin et al. (2023) and the "Just Ask" campaign launched by the RUG (University of Groningen, 2023). Hence, before gathering data, we did not know if our manipulations of harm salience and ambiguous psychological harm would work. For future replications of this study, we therefore strongly advise to pre-test the vignettes and harm-salience manipulations. It would be conceivable to design many vignettes and then group them according to their pre-tested harmfulness. This would allow for a design with different levels of harm contained in the vignettes. Furthermore, these vignettes should then also describe a wider range of situations. We acknowledge that the present study only covers two examples of non-physical, potentially harmful interactions. As Haslam and colleagues' (2020) discussions of concept-creep suggest, the harm-related concepts affected by concept creep are manifold, and depending on own experience or gender, different situations can be perceived as more or less harmful (Schein & Gray, 2018). For example, we suggest testing vignettes that describe subtle cases of racial discrimination, sexism or fat shaming.

Secondly, both of our examples of potentially harmful interactions were presented as instant messenger chats. Although this helped controlling for the influence of appearance, gender or other characteristics of the actors, the format also has disadvantages. An extensive body of research proposes that computer-mediated communication (CMC) is less suitable for conveying

feelings and attitudes than face-to-face communication, as non-verbal cues are filtered out (Venter, 2019). This also means that different prerequisites and mechanisms might be decisive for empathy in digital communication compared to face-to-face communication (Collins et al., 2024). The perception of a situation could, therefore, be greatly altered if participants were to watch a video instead of reading text messages.

Thirdly, one decisive factor remains open in the design of our vignettes: the intention of the perpetrator. The dyadic offender-victim relationship of TDM (Schein & Gray, 2018) assumes an intentional perpetrator, and previous research has shown that intentional harm is more strongly associated with moral disapproval than unintentional harm (Bleske-Rechek et al., 2023; Yoo & Smetana, 2019). It is not clear from our vignettes whether the perpetrators act intentionally, and we did not ask participants how they assess the intention. However, this could significantly influence the evaluation of harmfulness and thus also moral judgment. Future studies could compare vignettes in which the harm is intentional with unintentional vignettes to further explore this point, or measure participants assumptions of intent.

Another important limitation of our research is that we used a self-report measure to capture individual difference in empathetic concern. Self-report measures can be subject to biases, such as the social desirability bias. Since empathy is generally seen as a socially desirable trait, individuals could be drawn to overstating their own concern for others' well-being. In practice, this means that the high levels of EC found in our sample could reflect true characteristics of our participants or a general trend to overreport own EC.

However, if we assume that the high levels of EC are not influenced by social desirability bias but accurate, then the consistently high EC levels in our samples reveal another limitation. The predicted interaction effect suggests that the effect of the campaign would be particularly

pronounced in low EC individuals. Since low EC is largely underrepresented in our sample, future research should specifically target this population. This could also shed light on whether harm awareness interventions (e.g. social safety campaigns) basically serve to increase empathetic concern and are therefore particularly suitable for low EC populations. This argument is echoed by scholars who advocate empathy and compassion focused interventions to strengthen safety and cooperation in global crises such as the Covid-19 pandemic (Grant et al., 2022). We see great potential in empathy-based interventions and consider empathetic concern to be a powerful force in not only recognizing harm but wanting to do alleviate it. Consequently, it would be interesting to link not only harm perception and moral judgment, but also behavioral intentions such as helping or political activism to EC in future research.

**Conclusion**

As we have explored in this paper, the concept of harm is evolving, with growing awareness of non-physical forms of harm permeating institutions and organizations. We argued that increased awareness might lead to greater sensitivity towards harm, which we regard as a positive trend of growing empathetic concern in modern society. However, a subjective understanding of harm can also foster moral disagreements and deepen societal divides. Thus, creating safe spaces should not only morally condemn harm but also promote tolerance and respectful coexistence amidst differing perceptions of harm. Balancing empathy with respect for diverse viewpoints is crucial for fostering a genuinely safe society.

## References

Amsterdam, U. van. (2024, January 12). *About the social safety awareness campaign*. University of Amsterdam. https://www.uva.nl/en/about-the-uva/about-the-university/social-safety/campaign/about-the-campaign.html

Ball, C. L., Smetana, J. G., & Sturge-Apple, M. L. (2017). Following My Head and My Heart: Integrating Preschoolers' Empathy, Theory of Mind, and Moral Judgments. *Child Development*, *88*(2), 597–611. https://doi.org/10.1111/cdev.12605

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

Boehm, C. (1982). The evolutionary development of morality as an effect of dominance behavior and conflict interference. *Journal of Social and Biological Systems*, *5*(4), 413–421. https://doi.org/10.1016/S0140-1750(82)92069-3

Bridgland, V. M. E., Jones, P. J., & Bellet, B. W. (2023). A Meta-Analysis of the Efficacy of Trigger Warnings, Content Warnings, and Content Notes. *Clinical Psychological Science*, 21677026231186625. https://doi.org/10.1177/21677026231186625

Collins, A. M., Warburton, W. A., Bussey, K., & Sweller, N. (2024). Factor Structure and Psychometric Properties of the Digital Communication Empathy Scale (DCES). *International Journal of Human-Computer Studies*, *183*, 103183. https://doi.org/10.1016/j.ijhcs.2023.103183

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

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

Decety, J., & Cowell, J. M. (2014). Friends or Foes: Is Empathy Necessary for Moral Behavior? *Perspectives on Psychological Science*, *9*(5), 525–537. https://doi.org/10.1177/1745691614545130

Filipovic, J. (2014, March 5). We've gone too far with "trigger warnings." *The Guardian*. https://www.theguardian.com/commentisfree/2014/mar/05/trigger-warnings-can-be-counterproductive

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

Grant, L., Reid, C., Buesseler, H., & Addiss, D. (2022). A compassion narrative for the sustainable development goals: Conscious and connected action. *The Lancet*, *400*(10345), 7–8. https://doi.org/10.1016/S0140-6736(22)01061-3

Gray, K., & Kubin, E. (2024). *Victimhood: The Most Powerful Force in Morality and Politics*. https://doi.org/10.31234/osf.io/2n9m5

Groningen, U. of. (2023, April 17). *Social Safety*. University of Groningen. https://www.rug.nl/about-ug/policy-and-strategy/social-safety/

Haidt, J., & Haslam, N. (2016, April 10). Campuses are places for open minds – not where debate is closed down. *The Observer*. https://www.theguardian.com/commentisfree/2016/apr/10/students-censorship-safe-places-platforming-free-speech

Haslam, N., Dakin, B. C., Fabiano, F., McGrath, M. J., Rhee, J., Vylomova, E., Weaving, M., & Wheeler, M. A. (2020). Harm inflation: Making sense of concept creep. *European Review of Social Psychology*, *31*(1), 254–286. https://doi.org/10.1080/10463283.2020.1796080

Kakal, F., Whipkey, K., & Cajegas, L. (2020). *Persuasive Storytelling: How Campaigning Can Shift Social Norms*. CARE Nederlands. https://www.carenederland.org/content/uploads/2021/09/Persuasive-Storytelling-How-Campaigning-Can-Shift-Social-Norms.pdf

Kesterton, A. J., & Cabral De Mello, M. (2010). Generating demand and community support for sexual and reproductive health services for young people: A review of the Literature and Programs. *Reproductive Health*, *7*(1), 25. https://doi.org/10.1186/1742-4755-7-25

Lahiri, S., Bingenheimer, J. B., Evans, W. D., Wang, Y., Dubey, P., & Snowden, B. (2021). Social Norms Change and Tobacco Use: A Protocol for a Systematic Review and Meta-Analysis of Interventions. *International Journal of Environmental Research and Public Health*, *18*(22), 12186. https://doi.org/10.3390/ijerph182212186

Liu, X., Qu, W., & Ge, Y. (2022). The nudging effect of social norms on drivers' yielding behaviour when turning corners. *Transportation Research Part F: Traffic Psychology and Behaviour*, *89*, 53–63. https://doi.org/10.1016/j.trf.2022.06.011

McDermott, R., & Zimbardo, P. G. (2006). The Psychological Consequences of Terrorist Alerts. In B. Bongar, L. M. Brown, L. E. Beutler, J. N. Breckenridge, & P. G. Zimbardo (Eds.), *Psychology of Terrorism* (pp. 357–370). Oxford University Press. https://doi.org/10.1093/med:psych/9780195172492.003.0023

McGrath, M. J., Randall-Dzerdz, K., Wheeler, M. A., Murphy, S., & Haslam, N. (2019). Concept creepers: Individual differences in harm-related concepts and their correlates. *Personality and Individual Differences*, *147*, 79–84. https://doi.org/10.1016/j.paid.2019.04.015

Quissell, K. (2022). What's in a Norm? Centering the Study of Moral Values in Scholarship on Norm Interactions. *International Studies Review*, *24*(4), viac049. https://doi.org/10.1093/isr/viac049

Rotterdam, U. of. (2024). *We create social safety together*. We create social safety together

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

Slavich, G. M. (2020). Social Safety Theory: A Biologically Based Evolutionary Perspective on Life Stress, Health, and Behavior. *Annual Review of Clinical Psychology*, *16*(1), 265–295. https://doi.org/10.1146/annurev-clinpsy-032816-045159

Slote, M. (2003, December 19). *Moral Sentimentalism* [Lecture]. British Society for Ethical Theory, Queen's University, Belfast. https://www.jstor.org/stable/27504292

Stueber, K. (2019). Empathy. In *The Stanford Encyclopedia of Philosophy* (Fall 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/empathy/

Venter, E. (2019). Challenges for meaningful interpersonal communication in a digital era. *HTS Teologiese Studies / Theological Studies*, *75*(1). https://doi.org/10.4102/hts.v75i1.5339

Warner, L. A., Cantrell, M., & Diaz, J. M. (2022). Social Norms for Behavior Change: A Synopsis: WC406/AEC745, 1/2022. *EDIS*, *2022*(1). https://doi.org/10.32473/edis-wc406-2022

Wispé, L. (1991). *The psychology of sympathy*. Plenum Press.

Yoo, H. N., & Smetana, J. G. (2019). Children's moral judgments about psychological harm: Links among harm salience, victims' vulnerability, and child sympathy. *Journal of Experimental Child Psychology*, *188*, 104655. https://doi.org/10.1016/j.jecp.2019.06.008

## Appendix A

**Items Moral Outrage Scale**

1. (Name's) actions make me angry.
2. (Name's) actions are morally wrong.
3. (Name's) actions upset me.

*adapted from* S*kitka, Bauman, & Mullen (2004)*

## Appendix B

**Items Empathetic Concern (EC)**

1. I often have tender, concerned feelings for people less fortunate than me.
2. Sometimes I don't feel very sorry for other people when they are having problems.
3. When I see someone being taken advantage of, I feel kind of protective towards them.
4. Other people's misfortunes do not usually disturb me a great deal.
5. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
6. I am often quite touched by things that I see happen.
7. I would describe myself as a pretty soft-hearted person.

*derived from Davis (1983)*