



Do individualizing moral foundations protect from higher moral disengagement in imprisoned men and women?

Maria Campos-Rams^{1,2} · Mariola Paruzel-Czachura³ · Maria Luisa Vecina¹

Accepted: 23 October 2023
© The Author(s) 2023

Abstract

We aimed to analyze whether the individualizing moral foundations play a protective role against moral disengagement in a sample of 367 women and men in prison, and whether, in addition, moral foundations promote the intention to change the behaviors that led them to prison. Controlling for gender differences, we found support for the hypothesized connection between the individualizing moral foundations, as proposed by the Moral Foundations Theory (i.e., care and fairness), and the vast majority of moral disengagement mechanisms, especially with the two cruelest (attribution of blame and dehumanization). Regarding the intention to change the immoral behavior that brought our participants into prison, we found that, as hypothesized, the individualizing moral foundations positively and victim locus category negatively showed significant predictive power. This particular result could be especially relevant to improve the re-education interventions formulated for people in prison.

Keywords Care · Fairness · Attribution of blame · Dehumanization · Intention to change · Prisoners · Gender differences

People who break the law may have varying psychological profiles, and very different circumstances might bring them to prison. However, they may have something in common: Somehow, at some point in their lives, they disengaged from very important moral standards, especially those that protect people from being harmed and cheated. Care (or not harming) and fairness (or not cheating) have been traditionally considered the core of morality (Gilligan, 1982; Graham et al., 2011; Kohlberg, 1973, 1994; Sauer, 2015; Turiel et al., 1987) and, therefore, regulated by all judicial systems (Cross, 2010; Fijnaut, 2017; Fletcher, 1998; Gardner & Anderson, 2014; Herring, 2021; Holmes, 2009; Marmor, 2012; Waldron, 2009).

Although morality was first defined as justice (Kohlberg, 1973) and subsequently as care (Gilligan, 1982), both are currently considered as two key foundations on which moral decisions are based on (Graham et al., 2011; Schein & Gray, 2018). Gender differences in morality reasoning development have been described and associated with differences in socialization (Gilligan, 1982). In this respect, women seem to base their moral judgments more on the welfare of others than men (Atari et al., 2020; Giordano et al., 2002; Graham et al., 2009; Paruzel-Czachura et al., 2023; Paruzel-Czachura & Blukacz, 2021).

The Moral Foundations Theory (MFT; Graham et al., 2018; Haidt, 2008) conceives care and fairness, and their respective opposites, i.e., harm and cheating, as two innate and universally available psychological systems for detecting and reacting to others' pain and suffering, and caring about issues related to equality, justice, and rights. They are called individualizing moral foundations and can be analyzed together and separately. To account for cultural and gender differences, the MFT also proposes adding other psychological systems or foundations focused on protecting group interests (Graham et al., 2013; Haidt & Joseph, 2008). They are called binding moral foundations and address issues such as loyalty, authority, and purity.

✉ Maria Campos-Rams
macamp08@ucm.es

¹ Faculty of Psychology, Complutense University of Madrid, Campus de Somosaguas, Ctra. De Húmera, s/n, Pozuelo de Alarcón, Madrid 28223, Spain

² Institute of Psychology, University of Silesia in Katowice, Katowice, Poland

³ Penn Center for Neuroaesthetics, University of Pennsylvania, Pennsylvania, US

For the even more recent Theory of Dyadic Morality (TDM; Schein & Gray, 2018), care alone is the fundamental basis on which moral judgments are made and maintained. According to the TDM, harm or intentionally caused suffering involves two perceived and causally connected elements, that is, an intentional agent causing harm to a vulnerable patient. Although the TDM does not agree to any other moral system or foundation apart from care, it allows for different manifestations of care depending on contexts and cultures. In this respect, fairness could be considered a very close manifestation since treating others unfairly clearly involves harming them.

Can the Moral Disengagement, especially the attribution of blame and dehumanization, be understood as a deficit in Care and Fairness?

Moral disengagement originally refers to the process of reconstruing moral judgments with the result that the moral evaluation of behavior is changed (Bandura, 1999, 2016, 2018). However, the concept can also be studied as a dispositional construct, that is, the propensity toward moral disengagement (Moore, 2015). Process moral disengagement helps us to understand how people do harm and live with themselves (Bandura, 2016), while the propensity to disengage morally helps us to understand why some people engage more easily in unethical or immoral behaviors than others. Because studying moral disengagement as a process permits analyzing the variables that trigger it and the outcomes it leads to (Moore, 2015), in this study, we examined the process of moral disengagement in a sample of people that, judging by their prison sentence, may be also predisposed to moral disengagement.

It has been proposed that individuals may choose different mechanisms to be morally disengaged depending on the context. The eight mechanisms originally proposed by Bandura (1999, 2016, 2018) were grouped by him into four categories (i.e., agency, behavior, outcome, and victim locus). The agency category includes two mechanisms: displacement and diffusion of responsibility (Bandura, 2016, 2018). They allow people to absolve themselves of all responsibility (e.g., “somebody pressured me to do it”). The importance of denial of responsibility as a technique that paves the way for disengagement from social norms without explicitly confronting moral standards is subtle and was noted early (Sykes & Matza, 1957).

The behavior category includes three mechanisms: moral justification, euphemistic labeling, and advantageous comparison (Bandura, 2016, 2018). Through them, people can reorganize and redefine immoral behavior (e.g., robbery)

and understand it from a more positive angle (e.g., wealth redistribution). This cognitive adjustment of the notion of the behavior itself would explain how generally people justify their behavior by referring to a different moral principle, using names to change the appearance of the behavior, and comparing their actions with others to make them look irrelevant (Schaefer & Bouwmeester, 2020).

The outcome category includes only disregarding the consequences (Bandura, 2016, 2018). This mechanism permits misconstruing the effects of the actions by minimizing or ignoring the detrimental effect of the behavior, arguing that it was not that bad (e.g., “he/she does not really mind”). People who break the law may claim that their behavior did not really harm other people as a technique of denial of the injury (Sykes & Matza, 1957).

All these mechanisms can undoubtedly help justify harming others; however, the ones described in the fourth category (i.e., victim locus) involve a perceptual alteration towards humanity regarding their right to protection against harm (Bandura, 1999, 2016, 2018). The two mechanisms included, dehumanization and attribution of blame, have been singled out as especially despicable because they allow extremely violent behaviors (Bandura, 2016; Schaefer & Bouwmeester, 2020), usually condemned by law. These psychological adjustments permit to perceive victims as complicit, causing or deserving of immoral behavior (Schaefer & Bouwmeester, 2020), or even worse, as unworthy of human moral status (Tsang, 2002), which makes it impossible to empathize with their suffering.

This precise meaning concerning victim dehumanization and attribution of blame is gathered by the TDM and by the MFT when they consider that at least care, and its nearest manifestation, fairness, are the core of morality (Graham et al., 2018; Haidt et al., 2009; Schein & Gray, 2018). Protecting individuals’ rights not to be physically or psychologically harmed is deeply rooted in human nature and connected with both the capacity to empathize with the suffering of the vulnerable (Gilligan, 1982; Gilligan & Attanucci, 1988; Kohlberg, 1973, 1994; Sauer, 2015), and the need for reciprocal interactions, where cooperation and not selfish competition is the adaptive behavior (Graham et al., 2011; Haidt, 2008; Haidt & Graham, 2007; Trivers, 1971; von Hippel & Trivers, 2011).

Although there are more recent taxonomies (Schaefer & Bouwmeester, 2020), in this study, we analyzed our data in the light of Bandura’s (2016, 2018) because it includes together dehumanizing and blaming the victim. These two mechanisms can easily bring out the worst in people, and the worst implies harming someone. In this respect and since our participants were judged and condemned for serious crimes, we consider that the victim locus category could be highly represented. Nonetheless, our Supplementary

Materials will show results using the taxonomy recently proposed by Schaefer and Bouwmeester (2020).

Using conventional samples, some empirical evidence shows that individualizing moral foundations inhibit different violent behaviors (Silver & Silver, 2021) and promote reciprocal social behaviors (Graham et al., 2011; Haidt, 2008; Haidt & Graham, 2007). Parallel, several studies have shown that moral disengagement plays a large role in criminal activity (Delisi et al., 2013), cheating (Fida et al., 2022), antisocial behavior (Nasaescu et al., 2023), aggressive behavior (Gini et al., 2014), and false allegations in court setting (Clemente et al., 2019). However, little is known about the antecedents of moral disengagement (Hyde et al., 2010), especially in samples that have committed serious crimes, which would be crucial to treat and prevent disconnection between moral standards and behaviors. Even less is known about the particular relationships between the individualizing moral foundations and the moral disengagement mechanisms, which would aid a more comprehensive approach to serious immoral behaviors.

The current research

We aimed to analyze whether the individualizing moral foundations (care and fairness) play a protective role against moral disengagement in a sample of 367 women and men in prison, and whether, in addition, they promote the intention to change the behaviors that led them to prison. Since all our participants were serving a sentence in prison for irrefutable crimes such as murder, robbery, grievous bodily harm, violence against a partner, drug dealing, and so on, our rationale was that they disengaged from the individualizing moral system and that may be reflecting on the various forms of moral disengagement, especially on victim locus, the category associated to the most despicable and extreme violent behaviors.

Generally, we anticipated that care and fairness, currently considered as the most important moral foundations to protect the most sacred human right to life (Gilligan & Attanucci, 1988; Graham et al., 2011; Kohlberg, 1969, 1994; Sauer, 2015; Schein & Gray, 2018), will play a protective role against any kind of moral disengagement, but specifically, regarding the one which strips victims of their dignity, thus allowing the worst version of harm upon them (victim locus). Furthermore, we considered that this protective role should prevail regarding the intention to change the immoral behaviors that brought our participants into prison.

With these ideas, we expected to capture the process of moral disengagement as a state variable (Schaefer & Bouwmeester, 2020), knowing that “states are conceptualized as temporary conditions of mentality or mood, transitory

levels of arousal or drive” (Messick, 1987, p. 6). Given the characteristics of our sample, this state is particularly important and can reflect how people in prison reconcile their condemning of past behavior with their future behavior. That may contribute to our theoretical understanding of the connections between well-known moral theories and their applicability to prevent and treat problematic behaviors where most needed.

To empirically test these ideas and consider the importance of gender differences in morality (Gilligan, 1982), we also analyzed differences between women and men on the importance of individualizing moral foundations and the four categories of moral disengagement (i.e., behavior, agency, outcome, and victim locus) proposed by Bandura (1999). In this respect, a consistent body of research has shown that women have a greater endorsement of care and fairness (e.g., Atari et al., 2020; Graham et al., 2009, 2011), higher empathy for others (e.g., Baez et al., 2017; Toussaint & Webb, 2005), higher self-importance of moral identity (Paruzel-Czachura & Blukacz, 2021), and less disruptive behaviors (e.g., Heiskanen & Lietonen, 2016; Knežević, 2018; Steffensmeier & Allan, 1996) than men. In sum, we tested three specific hypotheses:

Hypothesis 1 There will be statistically significant differences between women and men in all the variables: women will present higher scores in care and fairness, lower in moral disengagement, and higher intention to change (i.e., not commit the crime again) compared to men.

Hypothesis 2 There will be negative relationships between the individualizing moral foundations and the eight moral disengagement mechanisms. Additionally, the individualizing moral foundations will negatively predict the worst of them, dehumanization and attribution of blame, grouped by Bandura in the victim locus category.

Hypothesis 3 The individualizing moral foundations will positively predict the intention to change the immoral behavior that brought our participants to prison, while at least the victim locus category will predict it negatively.

The data for the current study are publicly available at Open Science Framework: https://osf.io/av7fj/?view_only=None.

Methods

Participants and Procedure

We entered five prisons in a region of Poland. Regardless of the type of crime, every prisoner was invited by prison

authorities to take part in the study voluntarily and without monetary compensation. Because of safety reasons, researchers could not enter the prison. Informed consent was obtained from all prisoners. The participants were informed about the study's aim, their anonymity, the right not to provide specific information, and the right to decline participation at any moment. Pen and paper questionnaires were used in non-transparent envelopes to maintain their anonymity.

The total number of prisoners who were invited to participate was 7696, the response rate was 4.9% (our final sample was $N=382$). Sixteen cases were eliminated because of missing data or for supplying the same response more than 10 times in a row. As a result, 367 prisoners were included in our statistical analyses.

We conducted sensitivity power analyses using G*Power to determine our sample size. For the t-test (Hypothesis 1), this implied 80% power to detect an effect of $d=0.269$ with an alpha of 0.05. For the four linear regression with two predictors (Hypothesis 2), this sample size resulted in 80% power to detect an effect size of $f^2=0.026$ with an alpha of 0.05. For the linear regression with six predictors (Hypothesis 3), this sample size resulted in 80% power to detect an effect size of $f^2=0.038$ with an alpha of 0.05.

Our sample consisted of 230 men and 137 women. The average age was 38.15 years old ($SD=10.90$). Most participants had a low level of education (primary education: 25.1%; secondary education: 20.7%; technical education: 23.7%; professional education: 24%; high school: 3%; university: 3.5%). The most committed crimes included stealing (19.1%, $n=70$), fraud (16.9%, $n=62$), robbery (8.4%, $n=31$), and fighting or beating (7.9%, $n=29$). Forty-five prisoners (12.3%) did not report the type of crime they committed; the majority reported more than one crime.

Instruments

Moral foundations Questionnaire

We used the Polish translation (Jarmakowski-Kostrzanowski & Jarmakowska-Kostrzanowska, 2016) of the care and fairness scales of the Moral Foundations Questionnaire (Graham et al., 2011). Specifically, the relevance and the judgment subscales were used to assess 12 items in total. The judgment scale assesses levels of agreement with specific and contextualized moral judgment statements (1 = *strongly disagree*; 6 = *strongly agree*). For example, "One of the worst things a person could do is to hurt a defenseless animal" for care. The Relevance subscale assesses (1 = *definitely not important*; 6 = *definitely important*) how relevant it is for each participant several foundation-related concerns, for instance, "Whether or not some people were treated differently from others" for fairness.

Moral disengagement

We used the Polish translation of Detert et al. (2008)'s moral disengagement scale (Chudzicka-Czupala & Baran, 2016). This 24-item questionnaire measures the levels of the eight mechanisms of moral disengagement: moral justification, euphemistic labeling, advantageous comparison, displacement of responsibility, diffusion of responsibility, disregarding or distorting the consequences, attribution of blame, and dehumanization (1 = *strongly disagree*; 5 = *strongly agree*). For example: "It is alright to lie to keep your friends out of trouble" for moral justification, and "People who are mistreated have usually done things to deserve it" for attribution of blame.

Intention to change behavior

In order to evaluate if the person was willing to change the behavior that had brought them into prison (i.e., a real crime judged as such), we used a direct question: "Are you willing to change the behaviors that have brought you to your present situation?" (0 = *I am not going to change at all*; 5 = *I am going to change completely*).

Results

Descriptive statistics for all measured variables and t-test results depending on gender differences are in Table 1. As expected in Hypothesis 1, women scored higher in care and fairness than men and lower in agency, behavior, outcome, and victim locus categories than men. More specifically, they scored lower in all the moral disengagement mechanisms, except on diffusion of responsibility. Also, as hypothesized, women had a higher intention to change their immoral behavior than men. According to Cohen's guidelines (1992), the effect sizes were medium for most of these differences and especially large for the victim locus category, including dehumanization and attribution of blame.

Partial correlations for all measured variables, controlling for gender, can be seen in Table 2. As expected in Hypothesis 2, the individualizing moral foundations, together and separately, correlated negatively with most of the moral disengagement mechanisms, including dehumanization and attribution of blame, in such a way that the more participants cared about the individualizing moral foundations, the less moral disengagement they reported. Only moral justification and diffusion of responsibility were not related to care and fairness.

To deeply analyze these negative correlations, four regression analyses were conducted with a competitive method using the four moral disengagement categories as dependent

Table 1 Cronbach's Alpha, Descriptive Statistic for men and Women, and T-test Differences, Including Cohen's *d* Effect Sizes for all Measured Variables

| | Men | | | Women | | <i>t</i> | Cohen's <i>d</i> |
|--------------------------------|----------|----------|-----------|----------|-----------|----------|------------------|
| | <i>α</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| Individualizing | 0.83 | 4.23 | 0.82 | 4.71 | 0.69 | 5.81*** | 0.63 |
| Care | 0.74 | 4.26 | 0.95 | 4.80 | 0.73 | 5.76*** | 0.62 |
| Fairness | 0.67 | 4.21 | 0.84 | 4.61 | 0.72 | 4.70*** | 0.51 |
| Behavior Locus | 0.79 | 2.88 | 0.74 | 2.34 | 0.73 | -6.82*** | -0.74 |
| Moral Justification | 0.65 | 3.13 | 0.95 | 2.48 | 0.93 | -6.42*** | -0.69 |
| Euphemistic Labeling | 0.54 | 2.87 | 0.89 | 2.39 | 0.82 | -5.11*** | -0.55 |
| Advantageous Comparison | 0.69 | 2.64 | 1.01 | 2.14 | 0.89 | -4.76*** | -0.51 |
| Agency Locus | 0.72 | 3.08 | 0.78 | 2.84 | 0.83 | -2.71** | -0.29 |
| Displacement of Responsibility | 0.68 | 2.98 | 0.93 | 2.58 | 0.98 | -3.93*** | -0.42 |
| Diffusion of Responsibility | 0.60 | 3.17 | 0.94 | 3.10 | 0.92 | -0.66 | -0.07 |
| Outcome Locus | 0.61 | 2.7 | 1.28 | 2.16 | 0.87 | -4.34*** | -0.47 |
| Victim Locus | 0.78 | 2.86 | 0.89 | 2.16 | 0.72 | -8.23*** | -0.89 |
| Attribution of Blame | 0.68 | 2.87 | 1.03 | 2.29 | 0.81 | -5.67*** | -0.61 |
| Dehumanization | 0.71 | 2.91 | 1.05 | 2.02 | 0.84 | -8.50*** | -0.92 |
| Intention to Change | | 3.90 | 1.43 | 4.69 | 0.60 | 6.1*** | 0.66 |

*** $p < .001$; ** $p < .01$; * $p < .05$

variables and care and fairness together as the independent ones. These results are presented in Table 3. They confirmed our Hypothesis 2 and showed that, along with gender, the individualizing moral foundations had a significant relationship with the victim locus category, explaining an important percentage of its variance (20%).

A final regression analysis was conducted using intention to change as a dependent variable and the individualizing moral foundations, behavior, agency, outcome, and victim locus categories as independent variables. Results are shown in Table 4. As expected, gender, the individualizing moral foundations, victim and agency locus categories were significant predictors of intention to change. This means that being women, having a high regard for care and fairness, and having less deactivated mechanisms that distort humanity and displace responsibility contributed to having more intention to change the immoral behavior that brought our participants to prison. These four variables explained 15% of the total variance.

Discussion

Human beings are motivated to care for others and treat them fairly (Gilligan, 1982; Haidt & Graham, 2007; Trivers, 1971; Turiel et al., 1987). However, this motivation does not always succeed, and some people break the social rules intended to facilitate peaceful coexistence among large numbers of people.

Moral disengagement has often been connected with unethical and immoral behaviors (Bandura, 2016; Kouchaki et al., 2018), including different criminal behaviors

(Clemente et al., 2019; Delisi et al., 2013; Detert et al., 2008; Fida et al., 2022; Gini et al., 2014; Moore, 2008; Nasaescu et al., 2023), but we still need to look for antecedents (Hyde et al., 2010) and factors that can lower its level, especially if we aspire to prevent such behaviors or decrease the probability of repeating them in the future (recidivism) (Moore, 2015). Studying samples of people who broke the law and, with that, very important moral conventions, may contribute to this final objective.

In this study, we presented a primarily descriptive picture of women and men in prison in terms of the relationships between (1) their regard for the individualizing moral foundations, care and fairness, as part of the MFT and, in essence, also of the TDM (Graham et al., 2011; Haidt, 2008; Haidt & Graham, 2007; Schein & Gray, 2018); (2) their level on the moral disengagement as originally grouped by Bandura (1999); and (3) their intention to change the problematic behaviors that brought them into prison. As a result, we found a consistent pattern of significant relationships between these measures, controlling for gender differences.

Particularly, we could observe that, as expected, men in our sample showed less regard for care and fairness as well as less intention to change their previous behaviors than women. At the same time, men were significantly more disengaged from all of the moral mechanisms measured except diffusion of responsibility, which was similar among women and men. These differences support two research lines, one focusing on values related to care, which seem to be greater in women (Atari et al., 2020; Graham et al., 2009, 2011), and the other focusing on criminal behaviors, which tend to be greater in men (Heiskanen & Lietonen, 2016; Knežević, 2018; Steffensmeier & Allan, 1996).

Table 2 Pearson's Correlations Controlling for Gender for all Measured Variables

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|-----------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|--------|---------|--------|---------|--------|---------|
| 1. Individualizing | | | | | | | | | | | | | | |
| 2. Care | 0.93** | | | | | | | | | | | | | |
| 3. Fairness | 0.92** | 0.70** | | | | | | | | | | | | |
| 4. Behavior Locus | -0.21** | -0.20** | -0.21** | | | | | | | | | | | |
| 5. Moral Justification | -0.01 | -0.004 | -0.008 | 0.76** | | | | | | | | | | |
| 6. Euphemistic Labeling | -0.22** | -0.19** | -0.22** | 0.82** | 0.45** | | | | | | | | | |
| 7. Advantageous Comparison | -0.30** | -0.27** | -0.28** | 0.81** | 0.37** | 0.55** | | | | | | | | |
| 8. Agency Locus | -0.06 | -0.07 | -0.03 | 0.58*** | 0.40*** | 0.45*** | 0.49*** | | | | | | | |
| 9. Displacement of Responsibility | -0.12* | -0.14** | -0.08 | 0.61** | 0.40** | 0.52** | 0.53** | 0.85*** | | | | | | |
| 10. Diffusion of Responsibility | 0.024 | 0.020 | 0.024 | 0.38** | 0.28** | 0.33** | 0.30** | 0.85*** | 0.44** | | | | | |
| 11. Outcome Locus | -0.32** | -0.32** | -0.27** | 0.53** | 0.28** | 0.40** | 0.57** | 0.34** | 0.39** | 0.19** | | | | |
| 12. Victim Locus | -0.24** | -0.24** | -0.20** | 0.66** | 0.47** | 0.55** | 0.57** | 0.47** | 0.48** | 0.32*** | 0.50** | | | |
| 13. Attribution of Blame | -0.22** | -0.22** | -0.19** | 0.62** | 0.42** | 0.47** | 0.58** | 0.48** | 0.48** | 0.37** | 0.54** | 0.86** | | |
| 14. Dehumanization | -0.19** | -0.19** | -0.15** | 0.52** | 0.38** | 0.48** | 0.40** | 0.34*** | 0.35** | 0.22** | 0.34** | 0.87** | 0.49** | |
| 15. Intention to Change | 0.19** | 0.18** | 0.17** | -0.01 | 0.04 | -0.001 | -0.06 | 0.07 | -0.05 | 0.07 | -0.06 | -0.15** | -0.07 | -0.19** |

*** $p < .001$; ** $p < .01$; * $p < .05$

Confirming these gender differences in our specific sample of people in prison not only contributes to the visibility of women (Collins, 2010; Olson et al., 2016) but also may inspire positive interventions to reduce criminality. In this respect, we want to point out how morality began to be studied before as a matter of justice (Kohlberg, 1973), than as a matter of care (Gilligan, 1982), and how currently the common minimum that receives the most consensus to be considered the basis of morality is the last one (Schein & Gray, 2018; Sauer, 2015), in which precisely women score higher than men (Giordano et al., 2002; Graham et al., 2011). Since socialization is a process open to change, our results indirectly support the idea that if men were socialized as women, their moral disengagement would probably be similar to that of women, that is, lower.

Controlling for these relevant differences, we found that the more the participants scored on the individualizing moral foundations, the less moral disengagement they reported and the more they were willing to change the particular behavior judged as immoral by specific laws in a studied Western country. Particularly, we found support for (1) the hypothesized connection between the individualizing moral foundations and the worst moral disengagement mechanisms, attribution of blame and dehumanization, grouped by Bandura (1999) in the victim locus category, and (2) the hypothesized connections between the intention to change the immoral behavior that brought our participants into prison and both the individualizing moral foundations and victim locus category, the first one positively and the second one negatively.

Beyond our hypotheses, the individualizing moral foundations were also negatively related to the outcome and behavior locus categories. Additional analyses showed similar relationships using the taxonomy recently proposed by Schaefer and Bouwmeester (2020) (see Supplementary Materials). In this respect, it can be concluded that the individualizing moral foundations negatively predicted the three categories: reconstruing morality, reconstruing agency, and linguistic.

These results indirectly support our research idea about the protective role of the individualizing moral foundations, care and fairness, against moral disengagement in a particular sample of people in prison who committed serious crimes that undoubtedly hurt others. These findings can be read in light of the relevance granted to the individualizing moral foundations by the MFT and also by the TDM. Both theories seem opposite, but they agree that care/harm is likely the most important, frequent, and universal moral consideration. They also agree that the most despicable behaviors are evaluated by the amount of harm they cause to others (Graham et al., 2013; Schein & Gray, 2016, 2018). Although for the MFT, there are five different and

Table 3 Regression Analysis Conducted with a Competitive Method Regressing the Individualizing Foundations on Behavior, Agency, Outcome, and Victim Locus Controlling for Participants' Gender

| | | β | t | 95% CI | Adjusted R^2 |
|----------------|-----------------|---------|----------|----------------|----------------|
| Behavior Locus | Gender | 0.443 | 5.45*** | 0.283, 0.603 | 0.15 |
| | Individualizing | -0.133 | -4.18*** | -0.195, -0.07 | |
| Agency Locus | Gender | 0.233 | 2.71** | 0.064, 0.402 | 0.02 |
| | Individualizing | -0.066 | -1.21 | 0.064, 0.402 | |
| Outcome Locus | Gender | 0.302 | 2.48* | 0.062, 0.542 | 0.15 |
| | Individualizing | -0.313 | -6.57*** | -0.407, -0.219 | |
| Victim Locus | Gender | 0.615 | 6.74*** | 0.435, 0.795 | 0.20 |
| | Individualizing | -0.168 | -4.70*** | -0.238, -0.097 | |

*** $p < .001$; ** $p < .01$; * $p < .05$

Table 4 Regression Analysis Conducted with a Competitive Method Regressing the Individualizing Foundations, Behavior, Agency, Outcome, and Victim Locus on Intention to Change, Controlling for Gender

| | β | t | 95% CI | Adjusted R^2 |
|-----------------|---------|----------|----------------|----------------|
| Gender | -0.521 | -3.78*** | -0.792, -0.250 | 0.15 |
| Individualizing | 0.156 | 2.98** | 0.053, 0.259 | |
| Victim Locus | -0.275 | -3.27** | -0.441, -0.110 | |
| Agency Locus | 0.252 | 2.94** | 0.084, 0.421 | |
| Behavior Locus | 0.112 | 1.50 | | |
| Outcome Locus | 0.048 | 0.804 | | |

*** $p < .001$; ** $p < .01$; * $p < .05$

independent moral foundations, and for the TDM, moral judgment is only grounded in care, our results seem to confirm that this common minimum was related to some kind of protection against the particular mechanisms of humanity distortion, which divide individuals between the ones inside the moral domain and the ones outside, allowing despicable and extremely violent behaviors against the ones outside (Bandura, 2016).

This knowledge is especially relevant for psychological treatments that encourage people penalized by the justice system to reconnect with shared and collective life. The findings are important for the preventive interventions designed for people at risk of social exclusion. Fostering higher consideration for the individualizing foundations might lead to self-sanctions that connect people with their moral standards. At the same time, reassuring an endorsement for care and fairness may promote reciprocal social connections and enhance the individual capacity to empathize with others. Additionally, it may encourage the intention to change what is somehow recognized to have been done wrong, preventing relapse in behavior.

Our study is not free from limitations. First, our participants constituted a convenience sample of prisoners who freely wanted to participate. Although we used a total sample of 367 women and men penalized by the justice system, it can not represent the entire prison population, especially those from different cultural backgrounds (Henrich et al., 2010). Second, we used a cross-sectional design based on correlations that does not allow us to test for cause-effect

relationships or processes properly. It is true that such a sensitive sample is not only difficult to gather but also to study with the needed control of the variables. In this respect, although we used intention to change as a proxy highly related to actual behavior (Ajzen, 1991; Kan & Fabrigar, 2017; Sheeran, 2002), we did not control such self-reported answers and did not test whether their intention was effectively related to their behavior. Since we were not allowed to access the sample for more than one moment by the prison authorities, it is not possible to conclude in terms of causality. So, we can neither properly affirm that a deficit of care and fairness causes excessive moral disengagement nor that a higher intention to change is caused by higher levels of individualizing moral foundations and lower levels of dehumanization. Third, our sample was very diverse with respect to the reported crimes. It can be said that all were serious, but there were not enough participants in each separate group to compare specific types of crimes. In this respect, it is clear that killing is not the same as stealing and sexual assault is not the same as cheating, and this can affect the hypothesized relationships since the amount of harm is very different in all these cases (Schaefer & Bouwmeester, 2020). Future studies could try to test some crime-related differences following these initial results, and if the higher the harm inside a particular crime, the lower the regard for care and fairness and the higher the moral disengagement.

These clear limitations cannot overshadow the value of this study conducted with women and men who broke very important moral rules and were therefore imprisoned.

Precisely because we all expect prison experience to be corrective and dissuasive, it is worth noting the descriptive knowledge of this study as a starting point. In this respect, we can conclude that, at least in our sample of this understudied population, the individualizing moral foundations, which is the common minimum for different theoretical approaches (Gilligan, 1982; Graham et al., 2011; Kohlberg, 1994; Schein & Gray, 2018; Turiel et al., 1987), were related to the vast majority of the moral disengagement mechanisms and especially with the worst that distort the victims humanity and give free rein to the most harmful behaviors (Bandura, 2016), that is, with attribution of blame and victim dehumanization. The individualizing moral foundations also seem to be related to the intention to change, an important variable considered a positive indicator in social reintegration treatments. From this initial conclusion and with the ultimate aim to prevent and change serious behaviors threatening our peaceful coexistence, longitudinal and experimental research can be conducted focusing on connections between individualizing moral foundations and moral disengagement mechanisms. Results brought by our study could be very valuable for future psychosocial interventions of re-socialization, especially regarding the fact that the prison population grows worldwide (Augsburger et al., 2022; Nuytiens & Christiaens, 2016; Swavola et al., 2016).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12144-023-05352-z>.

Funding Mariola Paruzel-Czachura received funding from the NAWA Bekker Programme “Moral thinking and unethical behavior” (PN/BEK/2020/1/00058/DEC/1).

Declarations

Ethical statements The study was accepted by the Ethical Committee of the University of Silesia in Poland, number: KEUS.15/04.2020.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open practices The data for the current study are publicly available at Open Science Framework: https://osf.io/av7fj/?view_only=None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright

holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T).
- Atari, M., Lai, M. H. C., & Dehghani, M. (2020). Sex differences in moral judgements across 67 countries. *Proceedings Biological Sciences*, 287(1937), 20201201. <https://doi.org/10.1098/rspb.2020.1201>.
- Augsburger, A., Neri, C., Bodenmann, P., Gravier, B., Jaquier, V., & Clair, C. (2022). Assessing incarcerated women's physical and mental health status and needs in a Swiss prison: A cross-sectional study. *Health & Justice*, 10(1), 8. <https://doi.org/10.1186/s40352-022-00171-z>.
- Baez, S., Flichtentrei, D., Prats, M., Mastandueno, R., García, A. M., Cetkovich, M., & Ibáñez, A. (2017). Men, women... who cares? A population-based study on sex differences and gender roles in empathy and moral cognition. *PLOS ONE*, 12(6), e0179336. <https://doi.org/10.1371/journal.pone.0179336>.
- Bandura, A. (1999). Moral Disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193–209. https://doi.org/10.1207/s15327957pspr0303_3.
- Bandura, A. (2016). *Moral disengagement: How people do harm and live with themselves*. Worth Publishers.
- Bandura, A. (2018). Toward a psychology of Human Agency: Pathways and reflections. *Perspectives on Psychological Science*, 13(2), 130–136. <https://doi.org/10.1177/1745691617699280>.
- Chudzicka-Czupala, A., & Baran, L. (2016). Skłonność do moralnego dystansowania się studentów a ich intencje nieuczciwego zachowania się i reakcji na nieuczciwość akademicką innych [Tendency to moral disengagement among students and their intention to behave unethically and reaction to academic dishonesty of others]. *Chowanna*, 1, 251–272.
- Clemente, M., Espinosa, P., & Padilla, D. (2019). Moral disengagement and willingness to behave unethically against ex-partner in a child custody dispute. *Plos One*, 14(3), e0213662. <https://doi.org/10.1371/journal.pone.0213662>.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037//0033-2909.112.1.155>.
- Collins, R. E. (2010). The effect of gender on violent and nonviolent recidivism: A meta-analysis. *Journal of Criminal Justice*, 38, 675–684. <https://doi.org/10.1016/j.jcrimjus.2010.04.041>.
- Cross, N. (2010). *Criminal Law & Criminal Justice: An introduction*. SAGE publications. <https://doi.org/10.4135/9781446288283>.
- Delisi, M., Peters, D., Dansby, T., Vaughn, M., Shook, J., & Hochstetler, A. (2013). Dynamics of Psychopathy and Moral Disengagement in the etiology of crime. *Youth Violence and Juvenile Justice*, 12, 295–314. <https://doi.org/10.1177/1541204013506919>.
- Detert, J., Treviño, L., & Sweitzer, V. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *The Journal of Applied Psychology*, 93, 374–391. <https://doi.org/10.1037/0021-9010.93.2.374>.
- Fida, R., Ghezzi, V., Paciello, M., Tramontano, C., Dentale, F., & Barbaranelli, C. (2022). The Implicit Component of Moral Disengagement: Applying the Relational responding Task to Investigate its relationship with cheating behavior. *Personality & Social Psychology Bulletin*, 48(1), 78–94. <https://doi.org/10.1177/0146167220984293>.
- Fijnaut, C. (2017). *Criminology and the Criminal Justice System*. Intersentia.

- Fletcher, G. P. (1998). *Basic concepts of Criminal Law*. Oxford University Press.
- Gardner, T. J., & Anderson, T. M. (2014). *Criminal Law*. Cengage Learning.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Harvard University Press.
- Gilligan, C., & Attanucci, J. (1988). Two Moral orientations: Gender differences and similarities. *Merrill-Palmer Quarterly*, 34(3), 223–237.
- Gini, G., Pozzoli, T., & Hymel, S. (2014). Moral disengagement among children and youth: A meta-analytic review of links to aggressive behavior. *Aggressive Behavior*, 40(1), 56–68. <https://doi.org/10.1002/ab.21502>.
- Giordano, P. C., Cernkovich, S. A., & Rudolph, J. L. (2002). Gender, crime, and Desistance: Toward a theory of Cognitive Transformation. *American Journal of Sociology*, 107, 990–1064. <https://doi.org/10.1086/343191>.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046. <https://doi.org/10.1037/a0015141>.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385. <https://doi.org/10.1037/a0021847>.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology*, 47, 55–130. Academic Press. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>.
- Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitsch, C., & Moijman, M. (2018). Moral foundations theory: On the advantages of moral pluralism over moral monism. *Atlas of moral psychology*, 211–222. The Guilford Press.
- Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, 3(1), 65–72. <https://doi.org/10.1111/j.1745-6916.2008.00063.x>.
- Haidt, J., & Graham, J. (2007). *When Morality opposes Justice: Conservatives have Moral intuitions that liberals May not recognize*. Social Science Research Network.
- Haidt, J., & Joseph, C. (2008). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The innate mind*, 3, 367–391. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195332834.003.0019>.
- Haidt, J., Graham, J., & Joseph, C. (2009). Above and below left–right: Ideological narratives and moral foundations. *Psychological Inquiry*, 20(2–3), 110–119. <https://doi.org/10.1080/10478400903028573>.
- Heiskanen, M., & Lietonen, A. (2016). Crime and gender: A study on how men and women are represented in International Crime statistics. *Universität Tübingen*. <https://doi.org/10.15496/publikation-24365>.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29–29. <https://doi.org/10.1038/466029a>.
- Herring, J. (2021). *Criminal Law*. Bloomsbury Publishing.
- Holmes, O. W. (2009). *The Common Law*. Harvard University Press.
- Hyde, L. W., Shaw, D. S., & Moilanen, K. L. (2010). Developmental precursors of moral disengagement and the role of moral disengagement in the development of antisocial behavior. *Journal of Abnormal Child Psychology*, 38(2), 197–209. <https://doi.org/10.1007/s10802-009-9358-5>.
- Jarmakowski-Kostrzanowska, T., & Jarmakowska-Kostrzanowska, L. (2016). Polska adaptacja Kwestionariusza kodów moralnych (MFQ-PL) [The Polish adaptation of moral foundation questionnaire (MFQ-PL)]. *Psychologia Społeczna*, 11, 489–508.
- Kan, M. P. H., & Fabrigar, L. R. (2017). Theory of Planned Behavior. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences*, 1–8. Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1191-1.
- Knežević, M. (2018). To go or not to go: Personality, behaviour and neurophysiology of impulse control in men and women. *Personality and Individual Differences*, 123, 21–26. <https://doi.org/10.1016/j.paid.2017.10.039>.
- Kohlberg, L. (1973). *Moral development*. McGraw-Hill Films.
- Kohlberg, L. (1994). *Kohlberg's original study of moral development*. Garland Publishing.
- Kouchaki, M., Dobson, K. S. H., Waytz, A., & Kteily, N. S. (2018). The link between self-dehumanization and immoral behavior. *Psychological Science*, 29, 1234–1246. <https://doi.org/10.1177/0956797618760784>.
- Marmor, A. (2012). *The Routledge Companion to Philosophy of Law*. Taylor & Francis.
- Messick, S. (1987). Validity. *ETS Research Report Series*, 1987(2), i–208. <https://doi.org/10.1002/j.2330-8516.1987.tb00244.x>.
- Moore, C. (2008). Moral disengagement in processes of Organizational Corruption. *Journal of Business Ethics*, 80(1), 129–139. <https://doi.org/10.1007/s10551-007-9447-8>.
- Moore, C. (2015). Moral disengagement. *Current Opinion in Psychology*, 6, 199–204. <https://doi.org/10.1016/j.copsyc.2015.07.018>.
- Nasaescu, E., Zych, I., Ortega-Ruiz, R., Farrington, D. P., & Llorent, V. J. (2023). Stability and change in longitudinal patterns of antisocial behaviors: The role of social and emotional competencies, empathy, and morality. *Current Psychology*, 42(14), 11980–11994. <https://doi.org/10.1007/s12144-021-02484-y>.
- Nuytens, A., & Christiaens, J. (2016). Female pathways to crime and prison: Challenging the (US) gendered pathways perspective. *European Journal of Criminology*, 13(2), 195–213. <https://doi.org/10.1177/1477370815608879>.
- Olson, D. E., Stalans, L. J., & Escobar, G. (2016). Comparing male and female prison releasees across risk factors and Postprison Recidivism. *Women & Criminal Justice*, 26(2), 122–144. <https://doi.org/10.1080/08974454.2015.1083930>.
- Paruzel-Czachura, M., & Blukacz, M. (2021). How relevant for you is to be a moral person? Polish validation of the self-importance of Moral Identity Scale. *Plos One*, 16, <https://doi.org/10.1371/journal.pone.0255386>. Article e0255386.
- Paruzel-Czachura, M., Blukacz, M., Vecina, M., & Jonason, P. (2023). Moral foundations and criminality: Comparing community members to prisoners and violent/non-violent offenders. *Psychology, Crime, & Law*, accepted for publication.
- Sauer, H. (2015). Can't we all disagree more constructively? Moral foundations, Moral reasoning, and political disagreement. *Neuroethics*, 8(2), 153–169. <https://doi.org/10.1007/s12152-015-9235-6>.
- Schaefer, U., & Bouwmeester, O. (2020). Reconceptualizing moral disengagement as a process: Transcending overly Liberal and overly Conservative practice in the field. *Journal of Business Ethics*, 172, 525–543. <https://doi.org/10.1007/s10551-020-04520-6>.
- Schein, C., & Gray, K. (2016). Moralization and harmfulification: The dyadic loop explains how the innocuous becomes harmful and wrong. *Psychological Inquiry*, 27, 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>.
- Sheeran, P. (2002). Intention—Behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, 12(1), 1–36. <https://doi.org/10.1080/14792772143000003>.
- Silver, J. R., & Silver, E. (2021). The Nature and Role of Morality in Offending: A Moral foundations Approach. *Journal of*

- Research in Crime and Delinquency*, 58(3), 343–380. <https://doi.org/10.1177/0022427820960201>.
- Steffensmeier, D., & Allan, E. (1996). Gender and crime: Toward a gendered theory of female Offending. *Annual Review of Sociology*, 22, 459–487. <https://doi.org/10.1146/annurev.soc.22.1.459>.
- Swavola, E., Riley, K., & Subramanian, R. (2016). *Overlooked: Women and Jails in an Era of Reform*.
- Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review*, 22, 664–670. <https://doi.org/10.2307/2089195>.
- Toussaint, L., & Webb, J. (2005). Gender differences in the Relationship between Empathy and Forgiveness. *The Journal of Social Psychology*, 145, 673–685. <https://doi.org/10.3200/SOCP.145.6.673-686>.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35–57. <https://doi.org/10.1086/406755>.
- Tsang, J. A. (2002). Moral rationalization and the Integration of Situational Factors and psychological processes in immoral behavior. *Review of General Psychology*, 6(1), 25–50. <https://doi.org/10.1037/1089-2680.6.1.25>.
- Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, functions, and vagaries. *The emergence of morality in young children*, 155–243. University of Chicago Press. <https://doi.org/1988-97609-004>.
- von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34, 1–16. <https://doi.org/10.1017/S0140525X10001354>.
- Waldron, J. (2009). Judges as moral reasoners. *International Journal of Constitutional Law*, 7(1), 2–24. <https://doi.org/10.1093/icon/mon035>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.