Journal of Research in Personality 109 (2024) 104450

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Full Length Article

Associations of aversive ('dark') traits and affiliative ('light') traits with moral-dilemma judgments: A preregistered exploratory analysis using the CNI model

Nyx L. Ng [a,*], Craig S. Neumann [b], Dillon M. Luke [a], Bertram Gawronski [a]

[a] University of Texas at Austin, USA
[b] University of North Texas, USA

ARTICLE INFO

Keywords:
Light triad
Dark tetrad
Sadism
Moral judgment

ABSTRACT

Despite people's capacity for both good and evil, scant research has jointly examined the relations of affiliative and aversive traits with moral-dilemma judgments. Using the CNI model of moral-dilemma judgments, this preregistered exploratory study examined associations of aversive traits (Dark Tetrad comprising Machiavellianism, narcissism, psychopathy, sadism) and affiliative traits (Light Triad comprising Kantianism, humanism, faith in humanity) with sensitivity to consequences (*C*), sensitivity to moral norms (*N*), and general preference for inaction versus action (*I*) in responses to moral dilemmas. Dark Tetrad and Light Triad total scores were negatively and positively associated with sensitivity to moral norms, respectively. Sadism was the only trait-level predictor of moral-dilemma judgments, positively predicting sensitivity to consequences and negatively predicting sensitivity to moral norms.

From Rousseau to Milgram, people have for centuries been fascinated by humankind's capacity for both good and evil. Rather than being binarily good or bad, psychological research suggests that people possess a complex constellation of 'dark' aversive traits and 'light' affiliative traits (Kaufman et al., 2019). Whereas 'dark' aversive traits characterize one's antisocial propensity to manipulate others and exhibit callous behavior (Hilbig et al., 2021; Moshagen et al., 2018), 'light' affiliative traits characterize one's prosocial and beneficent orientation toward others (Kaufman et al., 2019; Neumann et al., 2020). In consideration of recent debates about whether the label 'dark' is (1) sufficiently precise to capture the antagonistic, aversive characteristics it is intended to (Kay & Arrow, 2022) and (2) sensationalistic and potentially stigmatizing (Kay & Arrow, 2023), we will instead use the terms 'aversive traits' and 'affiliative traits' in place of 'dark traits' and 'light traits', respectively.

Aversive traits and affiliative traits are differentially associated with a spectrum of beliefs, behaviors, and psychosocial outcomes, including aggression (Blais et al., 2014; Dinić et al., 2021), selfish behavior (Deutchman & Sullivan, 2018), unethical consumer attitudes and workplace behavior (Egan et al., 2015; Ellen et al., 2021; O'Boyle et al., 2012), relationship preferences and behaviors (Jonason & Buss, 2012; Jonason et al., 2012; Sevi & Doğruyol, 2020; Sevi et al., 2020), life satisfaction (Kaufman et al., 2019), motives for achievement and self-enhancement (Kaufman et al., 2019), and compliance and risky decision-making during the COVID-19 pandemic (Doerfler et al., 2021; Grežo & Adamus, 2022). Given these traits' antisocial versus prosocial elements, much extant work on such traits has investigated morally relevant beliefs and behaviors (e.g., violence, infidelity) or explicitly suggests a link between these traits and individual morality (e.g., Campbell et al., 2009; Jonason et al., 2015; Zuo et al., 2016). The current research is a preregistered exploratory investigation of the associations of aversive traits and affiliative traits with moral-dilemma judgments. Specifically, we explored associations of aversive traits (as measured by the Short Dark Tetrad; Paulhus et al., 2021) and affiliative traits (as measured by the Light Triad Scale; Kaufman et al., 2019) with sensitivity to consequences for the greater good, sensitivity to moral norms pertaining to harm and care, and general action tendencies in responses to sacrificial moral dilemmas.

1. 'Dark' Aversive Traits and 'Light' Affiliative Traits

Inspired by the dark side of the Force in the Star Wars trilogy (D. L. Paulhus, personal communication, June 8, 2020), Paulhus and Williams (2002) coined the term *Dark Triad* to refer to three socially aversive sub-

* Corresponding author at: Department of Psychology, University of Texas at Austin, 108 E Dean Keeton A8000, Austin, TX 78712, USA.
E-mail address: nyx@utexas.edu (N.L. Ng).

https://doi.org/10.1016/j.jrp.2023.104450
Received 5 July 2023; Received in revised form 15 December 2023; Accepted 18 December 2023
Available online 19 December 2023

clinical personality traits: (1) *Machiavellianism* (i.e., a penchant for amoral, exploitative, strategic-calculating behavior), (2) *narcissism* (i.e., an exaggerated sense of grandiosity and superiority coupled with egocentric, self-promoting behavioral tendencies), and (3) *psychopathy* (i.e., a constellation of affective-interpersonal deficits and impulsive antisocial behavioral tendencies). Research on aversive traits has since expanded to include the construct of subclinical *sadism* (i.e., a proclivity to hurt others for pleasure or subjugation) to form the *Dark Tetrad* (Buckels et al., 2013; Johnson et al., 2019; Međedović & Petrović, 2015; Paulhus, 2014; Paulhus et al., 2021; Plouffe et al., 2019). Whereas research on aversive traits is abundant (Muris et al., 2017; Neumann et al., 2022), work on its positive, prosocial counterparts has only started to take shape. Kaufman and colleagues (2019) expanded on the Dark Triad questionnaires to identify potential indicators of prosocial, benevolent behavior that are the conceptual opposites of Machiavellianism, narcissism, and psychopathy. Conceptualized as the antithesis to the Dark Triad, the *Light Triad* comprises three traits: (1) *faith in humanity* (i.e., belief in the fundamental goodness of mankind), (2) *humanism* (i.e., valuing the dignity and worth of others), and (3) *Kantianism* (i.e., treating others as ends unto themselves).

Aversive and affiliative traits are conceptually overlapping but distinct. The four aversive traits of Machiavellianism, narcissism, psychopathy, and sadism are said to share some characteristics including callousness (Jones & Figueredo, 2013) and empathy deficits (Heym et al., 2019; Jonason et al., 2013; Pajevic et al., 2018; Paulhus, 2014). Yet, they have unique associations with an array of psychosocial factors. Research on the bifactor model of empathy, for instance, has found unique empathy-deficit profiles for each aversive trait, such that only psychopathy and sadism are negatively associated with cognitive empathy and all aversive traits except for sadism are negatively associated with affective empathy (Pajevic et al., 2018). Likewise, after partialing out their shared variance, psychopathy but not Machiavellianism and narcissism has been linked to deficits in understanding moral normativity (Kay & Saucier, 2020). Research also suggests that, beyond the original triad of Machiavellianism, narcissism, and psychopathy, sadism independently predicts minimization of intention to cause harm and causal responsibility in moral judgment (Trémolière & Djeriouat, 2016).

Although Kaufman et al. (2019) had, in their development of the Light Triad, used existing measures of Machiavellianism, narcissism, and psychopathy to identify conceptually opposing affiliative traits, they argue that the three affiliative traits comprising the Light Triad are not mere inversions of Machiavellianism, narcissism, and psychopathy. In other words, the absence of aversive traits does not necessitate the presence of affiliative traits and vice versa. On the contrary, aversive and affiliative traits have been found to be only moderately negatively correlated (Kaufman et al., 2019; Lukić & Živanović, 2021). Recent research also suggests that affiliative traits have unique correlates that are distinct from the correlates of aversive traits despite their overlap (e.g., Sevi et al., 2020; Sevi & Doğruyol, 2020). Prior work linking this constellation of personality traits to moral-dilemma judgments suggests that aversive traits are associated with a stronger preference for utilitarian over deontological judgments (Bartels & Pizarro, 2011; Dinić et al., 2021; Patil, 2015), whereas affiliative traits are associated with a weaker preference for utilitarian over deontological judgments (Kaufman et al., 2019; Neumann et al., 2020). Extant work on this topic, however, is encumbered by three issues.

First, while there is converging evidence that aversive traits are positively associated with utilitarian moral-dilemma judgments (i.e., willingness to cause harm to maximize welfare for the greater good), findings pertaining to unique associations of the different aversive traits are mixed. Although several studies suggest a positive association between psychopathy and utilitarian moral-dilemma judgments (e.g., Bartels & Pizarro, 2011; Dinić et al., 2021; Djeriouat & Trémolière, 2014; Patil, 2015; Zamora et al., 2022), including a meta-analytic review finding a small-to-moderate effect ($r = 0.26$; Marshall et al., 2018), others suggest no relation between the two (e.g., Cima et al., 2010; Glenn et al., 2009). Evidence regarding the associations of Machiavellianism and narcissism with moral-dilemma judgments is similarly mixed: whereas some studies suggest weak to no relations between these two aversive traits and utilitarian moral-dilemma judgments (Dinić et al., 2021; Djeriouat & Trémolière, 2014), others suggest a positive association (Bartels & Pizarro, 2011; Zamora et al., 2022), particularly when utilitarian choices in moral dilemmas are motivated by self-interest rather than the greater good (Zamora et al., 2022). Research on the unique contributions of sadism in predicting utilitarian choices has produced similarly conflicting findings. Whereas Karandikar et al. (2019) found that sadism did not account for any incremental variance over Machiavellianism, narcissism, and psychopathy, Dinić and colleagues (2021) found that sadism not only accounts for additional variance in utilitarian moral-dilemma judgments over the other three aversive traits, but also over empathy and HEXACO traits.

In addition to these mixed findings, a second issue with extant research has been its exclusive focus on the association of the composite affiliative trait factor with moral-dilemma judgments (Kaufman et al., 2019; Neumann et al., 2020). Given the incipiency of research on affiliative traits (i.e., the Light Triad), unique associations of the three affiliative traits with moral-dilemma judgments have not yet, to our knowledge, been investigated. Because the three affiliative traits reflect different facets of a beneficent, prosocial orientation towards others, they may be differentially linked to moral-dilemma judgments. The affiliative trait of Kantianism, for example, directly reflects endorsement of deontological principles, specifically Immanuel Kant's second categorical imperative that one ought not to treat others as means to an end (Kant, 1785). Kantianism should therefore be positively associated with deontological judgments. In contrast, it is less clear how a belief in the fundamental goodness of humans (i.e., faith in humanity) would relate to moral-dilemma judgments.

One final issue hindering past research is the reliance on a methodologically flawed paradigm, which undermines interpretation of extant findings. To investigate the influence of aversive and affiliative traits on moral-dilemma judgments, researchers have typically relied on scenarios mirroring the classic trolley problem. For example, in the switch dilemma, a runaway trolley is said to kill five workers unless a lever is pulled to redirect the trolley to another track, where it would kill only one worker (Foot, 1967). According to the principle of utilitarianism, pulling the lever to sacrifice the one worker would be morally acceptable because it maximizes overall wellbeing by saving the lives of five others. According to the principle of deontology, pulling the lever would be morally unacceptable because it violates a fundamental moral norm that one should not kill others, regardless of the consequences. Researchers using this paradigm have thus interpreted endorsements of action as reflecting a utilitarian choice and endorsements of inaction as reflecting a deontological choice. This interpretation, however, is problematic for two reasons.

First, traditional moral dilemmas pit consequences for the greater good against adherence to moral norms (e.g., the utilitarian choice of sacrificing a few to save more necessarily infringes on the deontic moral norm of *do not harm*), regarding them as mutually exclusive. Responses to such moral dilemmas thus reflect relative preferences for utilitarian over deontological judgment instead of absolute preferences for either, which is problematic given that the processes underlying the two types of judgments are theorized to be independent (Conway & Gawronski, 2013). By confounding utilitarian and deontological judgment in a forced binary choice, it remains unclear whether aversive and affiliative traits are associated with tendencies to form utilitarian judgments, tendencies to form deontological judgments, or both.

Second, the moral dilemmas used in past research confound moral codes and action preferences, such that utilitarian judgments are conflated with a preference for action (e.g., taking the action described in the dilemma will sacrifice a few to save more) and deontological judgments are conflated with a preference for inaction (i.e., not taking

the action described in the dilemma will ensure adherence to a moral norm; see Crone & Laham, 2017). Whether aversive and affiliative traits are associated with tendencies to form utilitarian or deontological moral judgments or general preferences for action versus inaction in moral dilemmas therefore remains unclear. Given that aversive and affiliative traits can be associated with drastically different behavioral inclinations (e.g., psychopathy is linked to poor impulse control whereas Machiavellianism is characterized by caution and strategic-calculating behavior; Szabó & Jones, 2019), this methodological confound obfuscates interpretation of extant results.

2. The CNI Model

One tool that can be used to address the limitations of the traditional moral-dilemma paradigm is the CNI model of moral-dilemma judgments (Gawronski et al., 2017). The CNI model is a multinomial model that was developed to disentangle sensitivity to consequences (i.e., the *C* parameter), sensitivity to moral norms of harm (i.e., the *N* parameter), and general preference for inaction versus action (i.e., the *I* parameter) in responses to moral dilemmas. The three factors underlying moral-dilemma judgments are quantified using responses to matching sets of moral dilemmas that differ on two levels: (1) consequences for the greater good (i.e., the focal action described in the dilemma produces benefits that are either (a) greater than the costs or (b) smaller than the costs) and (2) relevant moral norms (i.e., the focal action described in the dilemma is either (a) prohibited by a proscriptive norm or (b) prescribed by a prescriptive norm). An example dilemma in its four variants is presented in Table 1.

As depicted in Fig. 1, each of the three factors underlying moral-dilemma judgments is captured by a parameter characterized by a unique pattern of responding across the four dilemma variants. The CNI model's *C* parameter captures the extent to which participants' responses in moral dilemmas are sensitive to consequences such that they (1) support action when the benefits associated with action outweigh their costs and (2) support inaction when the costs associated with action outweigh the benefits (first row in Fig. 1). The *N* parameter specifically captures the extent to which participants' responses to moral dilemmas are sensitive to moral norms of harm such that they (1) support action when the action prevents proximal harm and (2) support inaction when the action causes proximal harm (second row in Fig. 1). Lastly, the CNI model's *I* parameter captures the extent to which participants' responses reflect a general preference for inaction versus action such that they generally support inaction (versus action) regardless of cost-benefit ratios and type of norm (third and fourth rows in Fig. 1).

It should be noted that the CNI model's *N* parameter is not presumed to be driven by conscious, explicit thoughts about specific moral norms. Instead, the parameter simply captures the difference in responses between the cases wherein the action either causes or prevents proximal harm. Thus, although the *C* parameter could be argued to reflect the general norm *always maximize the benefits* (Hennig & Hütter, 2020), the response pattern captured by the *C* parameter (first row in Fig. 1) is still distinct from the one captured by the CNI model's *N* parameter on sensitivity to moral norms (second row in Fig. 1). Similarly, although the response pattern captured by the *I* parameter may be argued to reflect the general norm *first, do no harm* (Baron & Goodwin, 2020, 2021), this response pattern is again distinct from the one captured by the *N* parameter. Specifically, adherence to the general norm *first, do no harm* would be reflected in a preference for inaction regardless of the moral-dilemma variant presented (third and fourth rows in Fig. 1). The CNI model disentangles sensitivity to consequences, sensitivity to moral norms, and general preference for inaction versus action in responses to moral dilemmas by quantifying these three distinct response patterns.

To summarize, the *C* parameter on sensitivity to consequences captures the extent to which response patterns cohere with the utilitarian principle of maximizing welfare for the greater good. The *N* parameter on sensitivity to moral norms captures the extent to which response patterns cohere with the deontological principles of not causing harm and of preventing harm. Lastly, the *I* parameter on general preference for inaction over action captures the extent to which people prefer not to act in moral dilemmas.

Table 1
Example of a moral dilemma involving either a proscriptive or a prescriptive norm where the benefits of action are either greater or smaller than the costs of action. Dilemmas adapted from Gawronski, Armstrong, Conway, Friesdorf, and Hütter (2017). Reprinted with permission.

Proscriptive Norm Prohibits Action		Prescriptive Norm Prescribes Action	
Benefits of Action Greater than Costs	Benefits of Action Smaller than Costs	Benefits of Action Greater than Costs	Benefits of Action Smaller than Costs
You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and deadly to seniors and children. The only medication that can effectively stop the virus from spreading has severe side-effects. Although the virus will not kill her, the student suffers from a chronic immune deficiency that will make her die from these side-effects. Would you give the student the medication?	You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and can cause severe stomach cramps. The only medication that can effectively stop the virus from spreading has severe side-effects. Although the virus will not kill her, the student suffers from a chronic immune deficiency that will make her die from these side-effects. Would you give the student the medication?	You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and can cause severe stomach cramps. The student suffers from a chronic immune deficiency that will make her die from the virus if she is not returned to her home country for special treatment. However, taking her out of quarantine involves a considerable risk that the virus will spread. Would you take the student out of quarantine to return her to her home country for treatment?	You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and deadly to seniors and children. The student suffers from a chronic immune deficiency that will make her die from the virus if she is not returned to her home country for special treatment. However, taking her out of quarantine involves a considerable risk that the virus will spread. Would you take the student out of quarantine to return her to her home country for treatment?

Research using the CNI model has provided nuanced insights into the relationship between moral-dilemma responses and a range of factors, including basic personality traits (Luke & Gawronski, 2022), political ideology (Luke & Gawronski, 2021a), and other individual-difference factors (Körner et al., 2020). One line of research that is directly relevant to the present study pertains to associations of psychopathy with moral-dilemma judgments. Several studies suggest that psychopathy is negatively associated with all three factors captured by the CNI model, in that individuals high in psychopathy are less sensitive to consequences, less sensitive to moral norms, and less action averse in their responses to moral dilemmas (Gawronski et al., 2017; Körner et al., 2020; Luke & Gawronski, 2021b; but see Luke et al., 2022).

3. The Current Research

Using the CNI model, the current preregistered study exploratorily investigated the interrelationships between aversive traits, affiliative traits, and moral-dilemma judgments. We examined these associations both on a superordinate aggregate level (i.e., using total composite scores) and a subordinate facet level (i.e., using the individual subscale

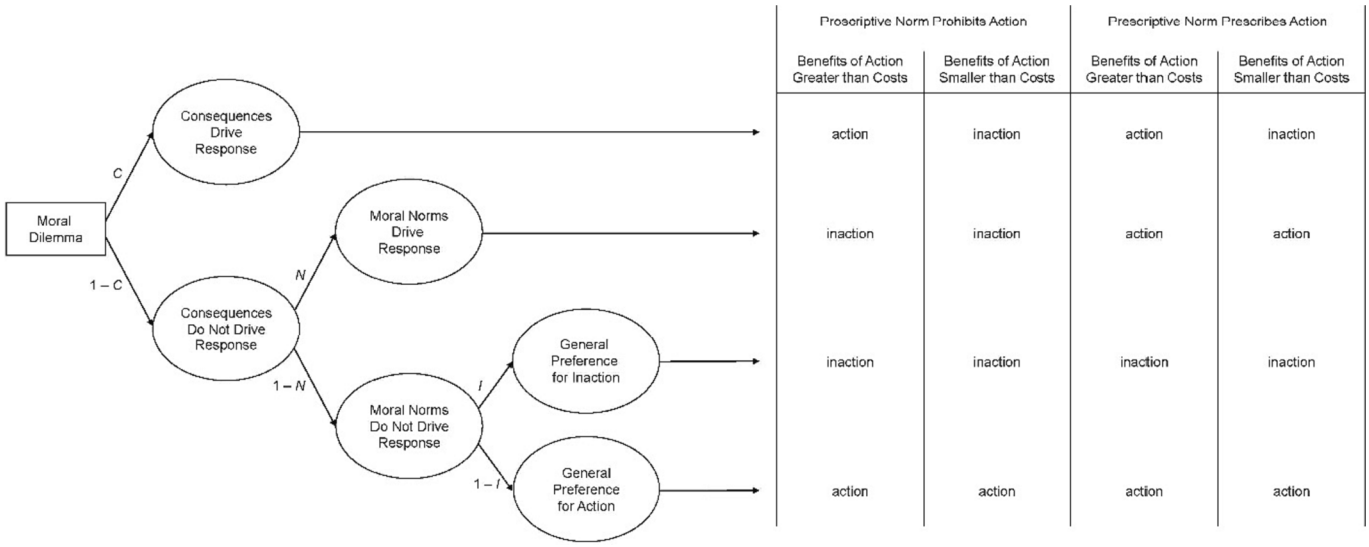

Fig. 1. CNI model of moral decision-making predicting action versus inaction responses in moral dilemmas with proscriptive and prescriptive norms and consequences involving benefits of action that are either greater or smaller than costs of action. *C* parameter = sensitivity to consequences; *N* parameter = sensitivity to moral norms; *I* parameter = general preference for inaction over action. Reproduced from Gawronski, Armstrong, Conway, Friesdorf, and Hütter (2017). Reprinted with permission from the American Psychological Association.

scores of Machiavellianism, narcissism, psychopathy, sadism, faith in humanity, humanism, Kantianism). The superordinate trait analyses were conducted to assess if an overarching aversive trait factor and affiliative trait factor are associated with moral-dilemma judgments, whereas the subordinate trait analyses examined which specific aversive and affiliative traits are associated with moral-dilemma judgments. The study was preregistered prior to data collection at https://osf.io/4apz2/. The data, analysis codes, and materials for the study can be accessed at https://osf.io/v3jxc/.

For comprehensiveness, we preregistered that we would test all potential associations between moral-dilemma judgments and the aversive and affiliative traits. Specifically, we examined whether the CNI model parameters for sensitivity to consequences, sensitivity to moral norms, and general preference for inaction versus action were each significantly associated with the superordinate aversive trait and affiliative trait. We also examined whether the three CNI model parameters were each significantly associated with the subordinate traits of Machiavellianism, narcissism, psychopathy, sadism, faith in humanity, humanism, and Kantianism.

4. Method

4.1. Participants

We preregistered to recruit a total of 850 participants, which was the largest sample feasible with the available funding. We expected approximately 15% of the sample to fail an attention check, leaving a sample of approximately 700 participants after preregistered exclusions. A sample of 700 participants would provide a power of 80% in detecting a correlation of $r = 0.11$ and a power of 95% in detecting a correlation of $r = 0.14$ with an alpha level of 0.05 (two-tailed). Participants were recruited on Prolific, a crowdsourcing platform for online research (Peer et al., 2017). Eligibility for participation was restricted to Prolific workers who (1) had the United Kingdom as their registered home country, (2) were at least 18 years of age, (3) were fluent in English, (4) had successfully completed at least 100 prior assignments, (5) had an approval rating of at least 95% across prior assignments, and (6) had not participated in a prior study from the authors' lab that used the same battery of moral dilemmas. Completion of the study took approximately 30 minutes, and participants were compensated $4.50 for their time. Because of accounting-related issues pertaining to the funding of the study, data collection was completed in two batches (i.e., 425 participants per batch), both of which were included in the sample for data analyses. Participants who terminated the study prior to completing all items did not receive compensation. Data from these participants were excluded from analyses. Following our preregistered exclusion criteria, participants with complete data were excluded from analyses if they failed an instructional attention check or provided the same response across dilemmas.[1]

A total of 852 participants completed all items.[2] Of these, 176 participants (20.7%) failed the attention check, resulting in a final sample of 676 participants. None provided the same response across dilemmas. Participant demographics are as follows: 67.5% female, 32.4% male, and 0.1% other; 91.4% identified as White, 4.4% Asian, 2.7% Black, 0.1% Hispanic/Latino/Spanish origin, 0.4% Middle Eastern/North African, and 1.9% other races; $M_{age} = 43.3$ years, $SD_{age} = 12.2$.

4.2. Measures

Short Dark Tetrad (SD4). The SD4 is a four-factor inventory that includes 28 items (seven items per construct) capturing *Machiavellianism*, *narcissism*, *psychopathy*, and *sadism* (Paulhus et al., 2021). Each item is rated on a five-point scale, ranging from *strongly disagree* (1) to *strongly agree* (5). The items used in the current research can be found in the Appendix of Paulhus et al. (2021). Responses to the SD4 were averaged to obtain the superordinate, total aversive trait score. Responses to each of the four SD4 subscales were averaged separately to obtain the subordinate mean psychopathy, narcissism, Machiavellianism, and sadism scores.

Light Triad Scale (LTS). The LTS is a 12-item inventory comprising three four-item scales measuring the light traits of *faith in humanity*, *humanism,* and *Kantianism* (Kaufman et al., 2019). Each item is rated on a five-point scale, ranging from *strongly disagree* (1) to *strongly agree* (5).

[1] Although it is possible that a participant chose the same response on all dilemmas based on careful consideration of the described scenarios, identical responses on all dilemmas more likely reflect lack of attention.

[2] We compensated two additional participants who were originally excluded from compensation. These participants completed all study items but did not submit their responses on the final page of the survey.

Responses to the LTS were averaged to obtain the superordinate, total affiliative trait score. Responses to each of the three LTS subscales were averaged separately to obtain the subordinate mean faith in humanity, humanism, and Kantianism scores.

CNI Model Moral-Dilemma Battery. The CNI model moral-dilemma battery includes 44 validated sacrificial dilemmas (Körner et al., 2020).[3] The dilemma battery comprised 11 basic scenarios, each with four variants manipulating (1) consequences for the greater good (i.e., the described action produces benefits that are either greater or smaller than the costs) and (2) moral norms (i.e., the described action is either prescribed or prohibited by a moral norm). Table 1 presents one of the 11 basic scenarios in its four variants. For each of the 44 dilemmas, participants indicated whether they would perform the described action using a binary *yes* (1) or *no* (0) answer choice. To analyze moral-dilemma responses, we summed the number of times participants chose to act in each of the four dilemma variants, providing four action indices per participant. To obtain indices of inaction responses, we subtracted the indices of action responses from 11 (i.e., the total number of dilemmas per variant), resulting in four inaction indices per participant. Both the action and inaction response indices can range from 0 to 11.

Using the action and inaction indices, we estimated the three CNI model parameters separately for each participant, following the procedures of Körner et al. (2020). Because the statistical underpinnings of the CNI model have been described in detail elsewhere (Gawronski et al., 2017; Körner et al., 2020), we only summarize some key aspects of the data aggregation here. Based on the processing tree depicted in Fig. 1, the CNI model provides four non-redundant equations that include the observed probability of action versus inaction responses on a given dilemma as known numerical values and the three model parameters as unknowns (see Gawronski et al., 2017). Using maximum-likelihood statistics, specific values for the three model parameters can be estimated, such that the discrepancy between the estimated probability of action (versus inaction) responses across dilemmas and the observed probability of action (versus inaction) responses across dilemmas is minimized.

In the current study, CNI model parameters were estimated for each participant by fitting the CNI model to the probabilities of action versus inaction responses across the four types of dilemmas (see Körner et al., 2020). Dependence of responses across dilemmas is accounted for by fitting the model at the level of individual participants. Following Gawronski et al. (2017), our analyses used a fixed estimation algorithm with random start values, two replications, and a maximum of 90,000 iterations. The CNI parameters were estimated with the freeware multiTree (Moshagen, 2010), and the template files for individual-difference research using the CNI model provided by Körner et al. (2020). To gauge the reliability of the three CNI model parameters in the current study, we estimated two scores for each parameter, one based on dilemmas with odd-item numbers and one based on dilemmas with even-item numbers (see Luke & Gawronski, 2022). The internal consistencies of the three parameters were estimated by calculating a Cronbach's α value for each parameter based on the two scores.

4.3. Procedure

After providing informed consent, participants completed the SD4 and LTS. The SD4 and LTS items were interspersed and presented in a fixed random order. Next, participants read and responded to the 44 moral dilemmas, which were also presented in a fixed random order. Afterward, participants responded to a set of demographic questions and an attention check. The attention check required participants to read a set of instructions, which asked participants not to answer a question (see Oppenheimer et al., 2009). Participants passed the attention check by *not* answering the question and moving on to the next page without selecting any answer choices. Finally, participants were provided with background information about the study and were redirected to Prolific for compensation.

4.4. Data Analytic Strategy

To investigate associations between aversive and affiliative traits with moral-dilemma judgments, we conducted a series of preregistered zero-order correlational, multiple-regression, and path analyses. In a first step, we analyzed the zero-order correlations between each of the four aversive traits, each of the three affiliative traits, the superordinate aversive and affiliative traits, and the CNI model parameter estimates. In a second step, we verified the robustness of the correlational results using multiple-regression analyses to control for shared variances between traits. Specifically, for the superordinate trait analyses, we entered the superordinate aversive and affiliative trait scores simultaneously as predictor variables in three separate models with the *C*, *N*, and *I* parameters as criterion variables. Likewise, for the subordinate trait analyses, we entered the four aversive traits (i.e., Machiavellianism, narcissism, psychopathy, and sadism) and three affiliative traits (i.e., faith in humanity, humanism, and Kantianism) simultaneously as predictor variables in three separate models with the *C*, *N*, and *I* parameters as criterion variables. In a third step, we conducted path analyses to investigate the associations between the study variables while controlling for the covariances between the CNI model parameters. These analyses served as an additional robustness check for the findings obtained in the correlation and multiple-regression analyses (see Supplemental Information for model fit results). Path analyses with maximum-likelihood estimation were conducted using M*plus* v8.0 (Muthén & Muthén, 2017).

5. Results

Table 2 details the descriptive statistics and internal consistencies of the study variables. The measures of aversive and affiliative traits showed no problematic deviations from normality. Overall, Cronbach's alphas for the SD4 and LTS measures suggest adequate internal consistencies, although mean inter-item correlation is arguably low for the SD4 total score (i.e., superordinate aversive trait) at 0.13. For the SD4 subscales (i.e., subordinate aversive traits), the Cronbach's alphas were borderline acceptable, ranging from 0.60 to 0.74, and the mean inter-item correlations ranged from 0.17 to 0.30. Of the four subscales, Machiavellianism had the lowest internal consistency. For the LTS subscales (i.e., subordinate affiliative traits), Cronbach's alphas ranged from 0.57 to 0.74 and mean inter-item correlations ranged from 0.26 to 0.43, with the Kantianism subscale having the lowest internal consistency. Reliability estimates for the CNI model parameters were similar to prior research using the CNI model (Luke & Gawronski, 2022), with the *C* and *N* parameters on sensitivity to consequences and moral norms having higher internal consistency than the *I* parameter on general preference for inaction over action.

5.1. Preregistered Analyses

Correlational Analyses. Table 3 presents the results of the zero-order correlational analyses. Regarding sensitivity to consequences, the correlational analyses revealed no significant associations with the superordinate aversive and affiliative traits. On a subordinate trait level, the analyses revealed a significant positive correlation between sensitivity to consequences and sadism. Sensitivity to moral norms showed a

[3] Körner et al.'s (2020) dilemma battery for research using the CNI model includes 12 basic scenarios in 4 different variants, summing up to a total of 48 dilemmas. However, an item-based analysis of these dilemmas revealed low construct validity of the moral-norms manipulation for one of the 12 basic dilemmas (Gawronski et al., 2020). To ensure high construct validity of our moral-judgment measure, this dilemma was not included in the current study.

Table 2
Means, standard errors, 95% confidence intervals, skewness, kurtosis, and internal consistency of the aversive traits, affiliative traits, four CNI model dilemma variants, and CNI model parameters.

Variables	Mean	*SE*	95% CI	Skewness	Kurtosis	α	Mean inter-item correlation
Aversive Trait	2.50	0.01	[2.47, 2.53]	0.35	0.02	0.81	0.13
Machiavellianism	3.41	0.02	[3.38, 3.45]	−0.19	0.29	0.60	0.17
Narcissism	2.45	0.02	[2.40, 2.49]	0.02	−0.30	0.74	0.30
Psychopathy	1.82	0.02	[1.78, 1.86]	1.02	1.65	0.69	0.25
Sadism	2.33	0.02	[2.29, 2.38]	0.83	0.84	0.69	0.24
Affiliative Trait	3.80	0.02	[3.76, 3.83]	−0.72	1.38	0.78	0.23
Faith in humanity	3.38	0.03	[3.32, 3.43]	−0.60	−0.01	0.74	0.43
Humanism	3.71	0.02	[3.67, 3.75]	−0.56	0.97	0.68	0.35
Kantianism	4.02	0.02	[3.98, 4.07]	−0.79	1.40	0.57	0.26
Proscriptive norm prohibits action							
Benefits > costs	3.59	0.08	[3.43, 3.76]	0.52	−0.13		
Benefits < costs	0.86	0.04	[0.77, 0.95]	1.68	3.20		
Prescriptive norm prescribes action							
Benefits > costs	9.21	0.06	[9.09, 9.33]	−1.00	0.64		
Benefits < costs	5.79	0.08	[5.63, 5.94]	0.23	−0.35		
CNI model parameters							
C parameter	0.28	0.01	[0.26, 0.29]	0.18	−0.62	0.58	0.45
N parameter	0.66	0.01	[0.64, 0.68]	−0.75	−0.30	0.57	0.41
I parameter	0.75	0.01	[0.73, 0.78]	−1.00	0.42	0.24	0.14

Note. Aversive trait = SD4 total score. Affiliative trait = LTS total score. *C* parameter = sensitivity to consequences; *N* parameter = sensitivity to moral norms; *I* parameter = general preference for inaction over action.

Table 3
Zero-order correlations between study variables.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Traditional score	1											
2. *C* parameter	0.73***	1										
3. *N* parameter	-0.62***	-0.09*	1									
4. *I* parameter	-0.29***	0.07	0.27***	1								
5. Aversive trait	0.20***	0.07	-0.21***	-0.06	1							
6. Affiliative trait	-0.08*	-0.00	0.18***	-0.01	-0.24***	1						
7. Machiavellianism	0.13***	0.04	-0.15***	-0.03	0.59***	-0.18***	1					
8. Narcissism	0.03	-0.02	-0.03	-0.03	0.67***	0.05	0.22***	1				
9. Psychopathy	0.14***	0.02	-0.15***	-0.04	0.75***	-0.24***	0.21***	0.38***	1			
10. Sadism	0.26***	0.14***	-0.25***	-0.05	0.73***	-0.31***	0.30***	0.19***	0.47***	1		
11. Faith in humanity	-0.04	0.01	0.11**	-0.05	-0.16***	0.82***	-0.14***	0.12**	-0.18***	-0.24***	1	
12. Humanism	-0.08*	-0.03	0.16***	-0.04	-0.15***	0.86***	-0.09*	0.15***	-0.18***	-0.29***	0.69***	1
13. Kantianism	-0.06	0.02	0.15***	0.06	-0.28***	0.67***	-0.23***	-0.18***	-0.18***	-0.18***	0.26***	0.35***

Note. *p <.05, **p <.01, ***p <.001. Aversive trait = SD4 total score. Affiliative trait = LTS total score. *C* parameter = sensitivity to consequences; *N* parameter = sensitivity to moral norms; *I* parameter = general preference for inaction over action.

significant negative correlation with the superordinate aversive trait and a significant positive correlation with the superordinate affiliative trait. On a subordinate trait level, sensitivity to moral norms had significant negative correlations with Machiavellianism, psychopathy, and sadism, and significant positive correlations with faith in humanity, humanism, and Kantianism. The correlational analyses revealed no significant associations between general preference for inaction versus action and the aversive and affiliative traits, both on a superordinate and subordinate factor level. Amongst the aversive and affiliative traits, the correlational analyses suggest an unexpected positive association between narcissism and both humanism and faith in humanity.

Regression Analyses. Table 4 presents the results of the multiple-regression analyses. Sensitivity to moral norms was negatively and positively predicted by the superordinate aversive and affiliative traits,

Table 4
Results of multiple-regression analyses regressing the CNI model parameters onto the aversive and affiliative traits.

	C parameter			*N* parameter			*I* parameter		
Variable	B	95% CI	Adj. R^2	B	95% CI	Adj. R^2	B	95% CI	Adj. R^2
Superordinate factors			0.00			0.06			0.00
Aversive trait	0.03†	[-0.00, 0.06]		−0.14***	[-0.19, −0.08]		−0.05	[-0.10, 0.01]	
Affiliative trait	0.01	[-0.02, 0.03]		0.08***	[0.04, 0.13]		−0.02	[-0.06, 0.03]	
Subordinate factors			0.02			0.07			0.00
Machiavellianism	0.01	[-0.02, 0.03]		−0.05†	[-0.09, 0.00]		−0.00	[-0.05, 0.05]	
Narcissism	−0.01	[-0.03, 0.01]		0.02	[-0.02, 0.06]		0.01	[-0.03, 0.05]	
Psychopathy	−0.01	[-0.04, 0.02]		−0.02	[-0.07, 0.02]		−0.01	[-0.06, 0.04]	
Sadism	0.05***	[0.02, 0.07]		−0.09***	[-0.13, −0.04]		−0.03	[-0.07, 0.02]	
Faith in humanity	0.02	[-0.01, 0.04]		−0.01	[-0.05, 0.03]		−0.02	[-0.06, 0.02]	
Humanism	−0.01	[-0.05, 0.02]		0.04	[-0.01, 0.10]		−0.03	[-0.09, 0.03]	
Kantianism	0.01	[-0.01, 0.04]		0.04†	[-0.00, 0.08]		0.04†	[-0.00, 0.08]	

Note. †p <.10, *p <.05, **p <.01, ***p <.001. Aversive trait = SD4 total score. Affiliative trait = LTS total score. *C* parameter = sensitivity to consequences; *N* parameter = sensitivity to moral norms; *I* parameter = general preference for inaction over action.

respectively. Distilling these associations into the subordinate aversive and affiliative trait factors, the multiple-regression analysis suggests that sadism largely accounts for the association between the superordinate aversive trait and sensitivity to moral norms, with Machiavellianism and Kantianism showing only marginal associations. Although the superordinate aversive trait did not significantly predict sensitivity to consequences, sadism emerged as a significant predictor in the subordinate traits model. Consistent with the correlational analysis, the regression analyses revealed no significant associations between general preference for inaction versus action and the aversive and affiliative traits, both on superordinate and subordinate trait levels.

Path Analyses. Two exploratory path models were specified with the CNI model parameters as criterion variables: one with the superordinate traits as predictors and another with the subordinate traits as predictors. As shown in Figs. 2 and 3, the results of the path analyses had the same patterns as the regression analyses. Specifically, even after controlling for the covariances between the three CNI model parameters, the superordinate aversive and affiliative trait factors were significantly associated with sensitivity to moral norms in a negative and positive direction, respectively. In addition, sadism also remained significantly associated with sensitivity to consequences and sensitivity to moral norms in a positive and negative direction, respectively.

5.2. Non-Preregistered Analyses

Traditional Score Analyses. One of the four dilemma variants included in the study structurally mirrors the traditional trolley paradigm, such that the focal action of the dilemma leads to greater benefits than costs but is prohibited by proscriptive norms of not inflicting harm. The action index for this dilemma variant can thus be used to compare our findings with past research that used dilemmas based on the trolley problem. Labelled as *traditional score,* higher scores on this index reflect a greater relative preference for utilitarian over deontological judgments. We analyzed the zero-order correlations of the traditional score with aversive and affiliative traits, respectively (Table 3). Consistent with past research, we found that the traditional score is positively correlated with the superordinate aversive trait and negatively correlated with the superordinate affiliative trait. On a subordinate trait level, the traditional score showed significant positive correlations with the aversive traits of Machiavellianism, psychopathy, and sadism, and a significant negative correlation with the affiliative trait of humanism.[4]

Exploratory Structural Equation Model. In the current study, we analyzed the associations between aversive and affiliative traits on both a superordinate (i.e., using total composite scores) and subordinate (i.e., using individual subscale scores) level. The superordinate trait analyses were conducted to assess if an overarching aversive trait factor and affiliative trait factor are associated with moral-dilemma judgments. There are empirical and clinical precedents for considering superordinate constructs with respect to psychopathy (Hare & Neumann, 2008; Neumann & Hare, 2008), as well as aversive traits in general (Bader et al., 2021), especially when these constructs are considered from a person-centered level. That is, persons with elevated scores on a specific aversive trait also tend to display heightened levels of other aversive traits. Similarly, research focusing on the development of a Light Triad construct has also provided empirical and clinical evidence for a superordinate affiliative trait factor (Kaufman et al., 2019; Neumann et al., 2020). In the current study, the Cronbach's alphas of the superordinate aversive and affiliative traits were acceptable ($\alpha s > 0.78$), and the correlations between each of the four aversive subscales ($rs > 0.19$) and between each of the three affiliative subscales ($rs > 0.26$) are significant and positive.

These results suggest that the superordinate factor analyses were appropriate. That being said, some may still object to our use of total composite measures (e.g., Glenn & Sellbom, 2015) for the following reasons: For one, if only some facets of the superordinate aversive and affiliative trait factors are related to moral-dilemma judgments, the superordinate trait analyses would yield weaker associations than it would have if all facets of the superordinate traits were equally associated with moral-dilemma judgments. One may also argue that the low mean inter-item correlation ($r = 0.13$) of the superordinate aversive trait suggests that this total composite measure may not reflect a unidimensional construct (Clark & Watson, 1995). In addition, the use of total composite measures may gloss over overlaps between aversive and affiliative subordinate traits, such as the positive association between the aversive trait of narcissism and the affiliative traits of humanism and faith in humanity. Our subordinate trait analyses resolve the first issue, as we modelled each of the seven subscales as separate predictors of moral-dilemma judgments. To address potential issues with the factor structure of the superordinate trait analyses, we re-analyzed the data using exploratory structural equation modelling (ESEM).[5]

An ESEM approach permits cross-loadings of subordinate aversive and affiliative traits onto both latent aversive and affiliative factors and therefore accounts for any overlaps between the traits, while also providing general internal construct validity for the latent aversive and affiliative latent domains. To conduct the ESEM, we freely loaded all seven traits onto two latent factors, which were then used to predict the CNI model parameters (Fig. 4). Results of this analysis suggest good model fit, CFI = 0.93, RMSEA = 0.07, 90% CI [0.06, 0.09]. In this ESEM model, the latent aversive factor was largely represented by psychopathy, followed by sadism and narcissism. Two affiliative traits significantly loaded onto this factor, but only one had a meaningful association. Specifically, Kantianism had a factor loading of -0.25 on the latent aversive factor, whereas humanism had an extremely trivial loading of 0.01 despite this relation reaching statistical significance. The cross-loading of Kantianism suggests that aversive tendencies are also linked to reduced concerns about treating people as ends unto themselves. The latent affiliative factor was largely represented by the traits of humanism and faith in humanity. Notably, with a factor loading of 0.31, the contribution of narcissism to the latent affiliative factor is comparable to that of the affiliative trait of Kantianism, which had a factor loading of 0.29. Sadism had the lowest factor loading for the latent affiliative factor at -0.15. The cross-loading of narcissism and sadism onto the latent affiliative factor suggests affiliative tendencies may be linked to positive self-views and to reduced propensities to enjoy harming others, respectively. In this model, the latent affiliative factor positively predicted sensitivity to moral norms, whereas the latent aversive factor negatively predicted it. The latent aversive factor had a marginal negative association with general preference for inaction over

[4] The identified associations between the traditional score and both the superordinate affiliative trait and psychopathy became non-significant after potential outliers were excluded (see Table S2 of Supplemental Information).

[5] Aside from using an ESEM approach to model the shared variances between correlated aversive and affiliative traits, another approach is to model a latent general factor accounting for the shared variances between all seven traits (e.g., see Horsten et al., 2023 for a bifactor model of aversive traits). This latent factor representing the shared characteristics of the seven aversive and affiliative traits, along with each trait's remaining unique effects, can then be used to predict the CNI model parameters. In one model, we loaded the subscales of the SD4 and LTS onto a latent general factor and then specified all factors (one latent and seven unique) as orthogonal. The model fit was poor, CFI = 0.61, RMSEA = 0.21. In another model, we loaded each item onto a latent general factor, loaded the aversive trait items onto a latent aversive factor, and loaded the affiliative trait items onto a latent affiliative factor. The three latent factors are then specified as orthogonal. However, model fit was still poor, CFI = 0.65, RMSEA = 0.07. In a third model, we loaded each item onto a latent general factor, loaded each item onto the seven specific factors, and then specified all factors as orthogonal. This improved model fit, but it was still suboptimal, CFI = 0.75, RMSEA = 0.06. Specifying the general factor as non-orthogonal to the seven specific factors in this model resulted in non-convergence.

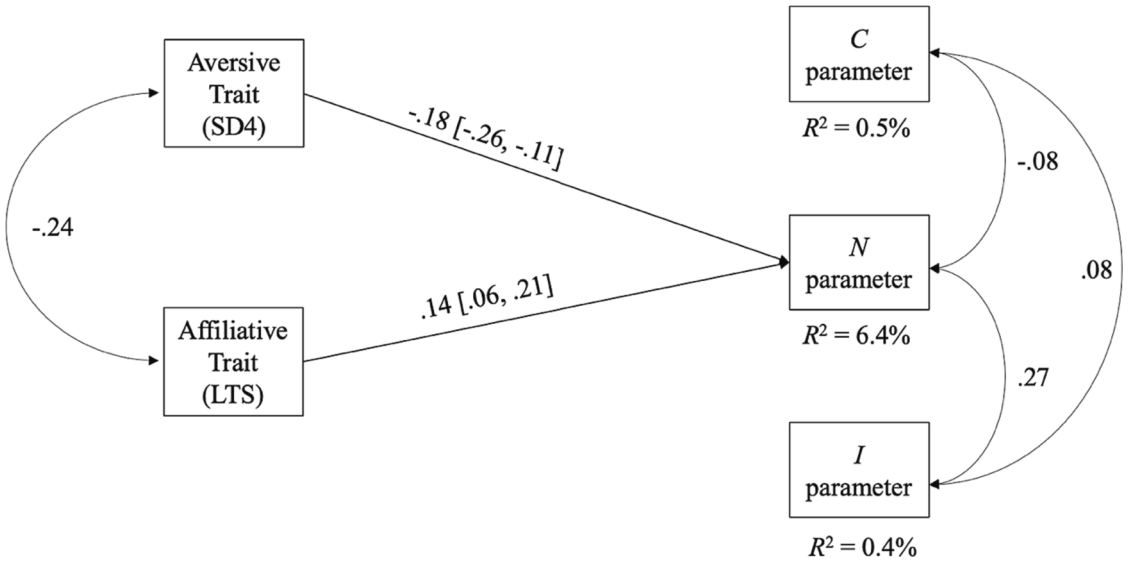

Fig. 2. Results of the path analysis with the superordinate aversive (SD4) and affiliative traits (LTS) as predictor variables and CNI model parameters as criterion variables (standardized parameters) with 95% confidence intervals in parentheses. Non-significant paths where $p \geq 0.05$ are omitted. *C* parameter = sensitivity to consequences; *N* parameter = sensitivity to moral norms; *I* parameter = general preference for inaction over action.

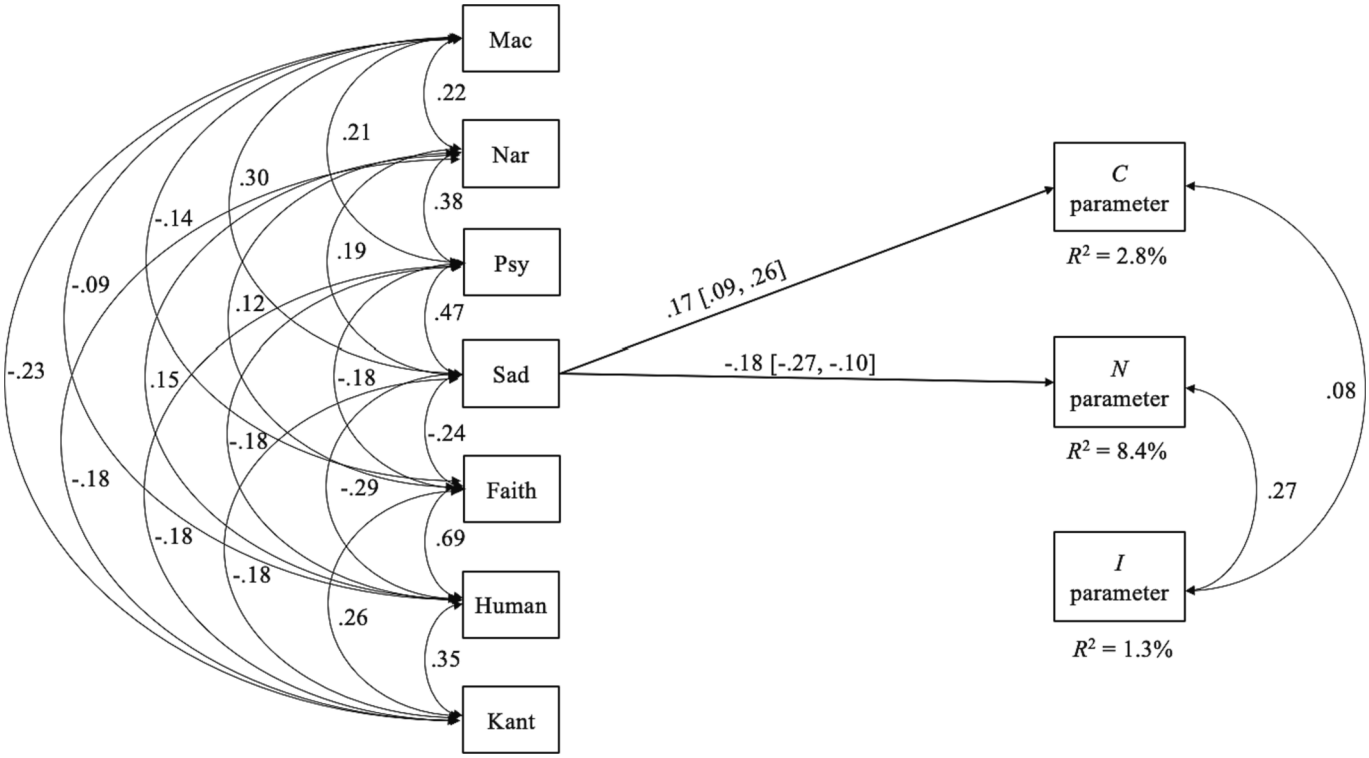

Fig. 3. Results of the path analysis with the subordinate aversive (SD4) and affiliative (LTS) traits as predictor variables and CNI model parameters as criterion variables (standardized parameters) with 95% confidence intervals in parentheses. Non-significant paths where $p \geq 0.05$ are omitted. Mac = Machiavellianism; Nar = Narcissism; Psy = Psychopathy; Sad = Sadism; Faith = Faith in humanity; Human = Humanism; Kant = Kantianism. *C* parameter = sensitivity to consequences; *N* parameter = sensitivity to moral norms; *I* parameter = general preference for inaction over action.

action β = -0.09, p =.07, 95% CI [-0.18, 0.01]. Thus, the results of the ESEM analyses are consistent with the findings of our analyses of the total composite scores of aversive and affiliative traits.

5.3. Supplemental Analyses

Outlier Analyses. To check the robustness of our findings, we re-ran all analyses after excluding potential outliers, which were identified as cases that laid beyond 2.5 absolute deviations from the median (Leys et al., 2013) for at least one of the seven trait measures. The results of the outlier analyses are reported fully in the Supplemental Information. We note only discrepancies with the preregistered multiple regression and path analyses here. Inconsistent with the full-sample analyses predicting sensitivity to moral norms, the previously marginal predictors of Machiavellianism and Kantianism reached statistical significance after outlier exclusion. However, it should be noted that excluding the

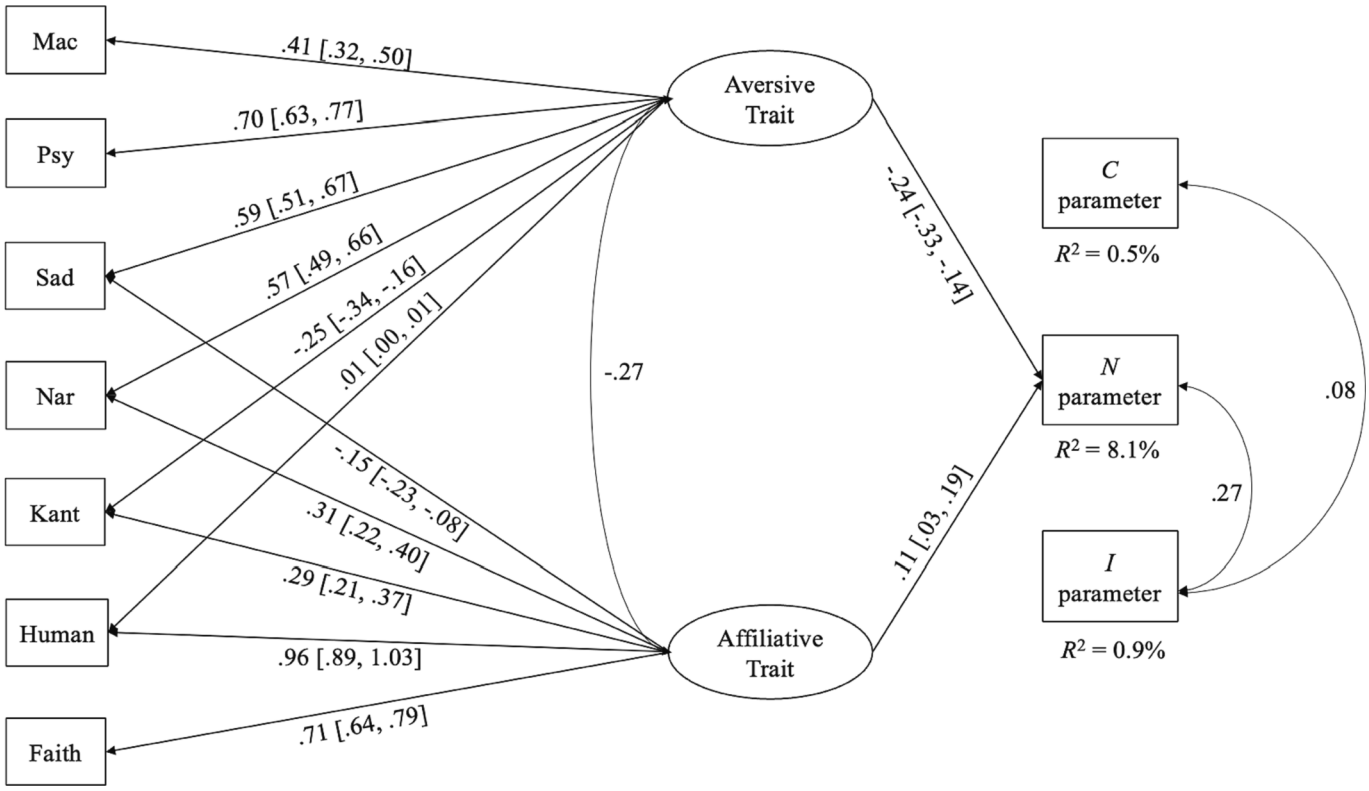

Fig. 4. Results of the ESEM model with cross-loaded latent aversive and affiliative factors predicting the CNI model parameters (standardized parameters) with 95% confidence intervals in parentheses. Non-significant paths where $p \geq 0.05$ are omitted. Mac = Machiavellianism; Nar = Narcissism; Psy = Psychopathy; Sad = Sadism; Faith = Faith in humanity; Human = Humanism; Kant = Kantianism. *C* parameter = sensitivity to consequences; *N* parameter = sensitivity to moral norms; *I* parameter = general preference for inaction over action.

potential outliers led to a reduction in internal consistency across the subscales, with both the subscales of Machiavellianism (α = 0.51, mean inter-item correlation = 0.13) and Kantianism (α = 0.21, mean inter-item correlation = 0.07) showing poor reliability after outlier exclusion. These findings should thus be interpreted with caution.

CAN Algorithm. In the CNI model, the *I* parameter on general preference for inaction over action is positioned at the bottom of the processing tree for methodological reasons (see Gawronski et al., 2017, 2020), which results in estimations of the *I* parameter having greater proportions of measurement error than the *C* and *N* parameters on sensitivity to consequences and moral norms. The null associations found between all traits and the CNI model's *I* parameter could have thus been an artefact of the CNI model's hierarchical structure. In anticipation of concerns about this issue, we conducted supplemental analyses using an alternative model called the CAN algorithm, which algebraically calculates the three moral judgment parameters concurrently (Liu & Liao, 2021). Full results of this set of re-analyses are reported in the Supplemental Information. We note only discrepant findings here.

Although the zero-order correlational and path analyses suggest a positive association between sadism and the CAN algorithm's *C* parameter on sensitivity to consequences, this association was non-significant in the multiple regression analyses that controlled for the high correlation between the CAN algorithm's parameters. Regarding the CAN algorithm's *N* parameter on sensitivity to moral norms, the regression analyses suggest both Machiavellianism and sadism are negative predictors whereas Kantianism is a positive predictor of sensitivity to moral norms. Yet, in the path analyses, only Machiavellianism and sadism emerged as significant predictors of sensitivity to moral norms. Crucially, the analyses also suggest a potential positive relation between affiliative traits and the CAN algorithm's *A* parameter, which captures one's preference to respond with action (versus inaction) in moral dilemmas. These results suggest that the low reliability of CNI model's *I* parameter may have suppressed associations with trait variables in the current study.

Latent Profile Analysis. In addition to the variable-centered analyses, we conducted an exploratory person-centered latent profile analysis of the LTS and SD4 scales. This set of analyses were preregistered with the goal of replicating the subtypes reported by Neumann et al. (2020). We replicated the findings of Neumann et al. (2020) and report the full results in the Supplemental Information.

6. Discussion

Because people possess a mélange of prosocial and antisocial characteristics (Kaufman et al., 2019; Neumann et al., 2020), extant research that solely investigated the associations of aversive traits with moral judgments has neglected one side of the picture. To date, research on the joint roles of aversive and affiliative traits in moral judgments remains scant. Moreover, extant work on the topic is limited by two methodological flaws that (1) confounded endorsements of moral doctrines with general action tendencies (Crone & Laham, 2017) and (2) framed deontological and utilitarian responses as binary opposites (Conway & Gawronski, 2013). Using the CNI model to resolve these limitations (Gawronski et al., 2017), the current research explored associations of aversive and affiliative traits with three aspects of moral-dilemma judgments, namely sensitivity to consequences, sensitivity to moral norms, and general preference for inaction versus action.

Assessing the traits on a superordinate factor level, neither the superordinate aversive trait nor affiliative trait were associated with sensitivity to consequences and general preference for inaction versus action. In contrast, the superordinate aversive trait factor negatively correlated with sensitivity to moral norms, whereas the superordinate affiliative trait factor positively correlated with sensitivity to moral

norms. These relationships remained statistically significant even when we (1) controlled for the covariance between the two superordinate factors, (2) controlled for the covariances between the three criterion variables, (3) used an ESEM approach to model latent affiliative and aversive factors, (4) excluded potential outliers (see Supplemental Information), and (5) used a different computational approach (i.e., the CAN algorithm; see Supplemental Information) to estimate the *N* parameter on sensitivity to moral norms.

Distilling the associations found in the superordinate trait analyses into the subordinate trait level, the preregistered correlational analyses revealed (1) a significant positive correlation between sensitivity to consequences and sadism, (2) significant negative correlations between sensitivity to moral norms and Machiavellianism, psychopathy, and sadism, and (3) significant positive correlations between sensitivity to moral norms and the three affiliative traits. None of the subordinate traits correlated with general preference for inaction versus action. Partialing out (1) the shared variance of the predictor variables in the multiple regression analyses and (2) the covariances between predictor and criterion variables in the path analyses, only two associations remained statistically significant: the positive association between sadism and sensitivity to consequences, and the negative association between sadism and sensitivity to moral norms. Comparatively, the negative association between sadism and sensitivity to moral norms appears more robust than the positive association between sadism and sensitivity to consequences. The negative association between sadism and sensitivity to moral norms emerged regardless of analytic approach, whether potential outliers were excluded, and whether the parameters were estimated using the CNI model or CAN algorithm. In contrast, the positive association between sensitivity to consequences and sadism did not reach statistical significance when sensitivity to consequences was estimated using the CAN algorithm and entered into a regression model that controlled for the high correlations between the moral-dilemma judgment parameters (see Supplemental Information).

6.1. Aversive 'Dark' Tetrad, Affiliative 'Light' Triad, and Moral-Dilemma Judgments

Our finding that the superordinate aversive trait is not associated with sensitivity to consequences seemingly contradicts prior research suggesting that individuals with elevated aversive traits tend to form more utilitarian judgments (e.g., Bartels & Pizarro, 2011; Djeriouat & Trémolière, 2014) whereas those with elevated affiliative traits tend to form fewer utilitarian judgments (Kaufman et al., 2019). However, because past studies had presumed that a rejection of a utilitarian stance would be equivalent to an endorsement of a deontological stance and vice versa, their findings could be interpreted in at least two distinct ways.

Specifically, past findings on the relationship between aversive traits and utilitarian responses to moral dilemmas could either be due to (1) a positive relation between aversive traits and utilitarian tendencies or (2) a negative relation between aversive traits and deontological tendencies. When we analyzed the correlations between the superordinate aversive and affiliative trait factors and responses to the traditional moral dilemma variant (i.e., dilemmas wherein the focal norm prescribes action and the benefits of acting outweigh the costs), we replicated the associations found in past research: aversive traits are positively correlated with relative preference for utilitarian over deontological responses, whereas affiliative traits are negatively correlated with it (although the latter relationship does not appear robust[4]). However, by manipulating both the levels of consequences and type of moral norm relevant to the moral dilemmas, we could tease apart utilitarian and deontological response patterns and clarify extant findings. Instead of conforming to the widely accepted interpretation that aversive and affiliative traits predict utilitarian tendencies, our research suggests that superordinate aversive and affiliative traits negatively and positively predict tendencies to adhere to moral norms instead of tendencies for welfare maximization, respectively.

6.2. Sadism and Moral-Dilemma Judgments

Because the CNI model's *N* parameter on sensitivity to moral norms reflects a behavioral tendency to conform to relevant moral norms in responses to moral dilemmas, and because aversive traits are associated with an array of norm-violating behaviors (Furnham et al., 2013; Paulhus & Williams, 2002), associations between aversive traits and sensitivity to moral norms, if significant, should trend towards the negative. Consistent with this idea, we find that the SD4 measure of sadism is negatively related to the extent to which people (1) show a preference for inaction when action causes proximal harm to a focal target and (2) show a preference for action when action prevents proximal harm to a focal target. Elevated sadism is essentially linked to norm-violating tendencies (i.e., choosing to act in scenarios that proscribe acting and choosing not to act in scenarios that prescribe acting).

Counter-normative tendencies aside, the current study also revealed a potential positive relationship between sadism and sensitivity to consequences. Because the CNI model's *C* parameter reflects a behavioral tendency to maximize overall outcomes for the greater good, and because aversive traits are characterized by tendencies toward self-interest at the expense of others' welfare (Zamora et al., 2022), associations between aversive traits and sensitivity to consequences should theoretically be negative in direction. Moreover, because sadism is characterized by enjoyment of others' suffering, one might even expect individuals with elevated levels of sadism to be more likely to endorse action when the benefits of the action are less than the costs and endorse inaction when the benefits of the action outweigh the costs. In other words, individuals who are highly sadistic might be expected to exhibit a harm-maximizing response pattern opposite to that of a welfare-maximizing response pattern. Yet, our results suggest the opposite: instead of harm-maximization, sadism is positively associated with welfare-maximization in moral dilemmas.

Why might sadism be linked to utilitarian, welfare-maximizing tendencies? Limitations of the SD4 measure of sadism could have contributed to this counterintuitive finding. A multifaceted trait, sadism can be delineated into vicarious (e.g., enjoyment of violent shows) and direct (e.g., enjoyment of inflicting harm) facets, as well as physical and verbal facets (Foulkes, 2019). However, not all facets of sadism are captured in the SD4 measure of sadism, which includes four items on vicarious physical sadism and three items on direct and vicarious verbal sadism. None concern direct physical sadism, which has been surmised to be a better predictor of antisocial tendencies than vicarious forms of sadism (Foulkes, 2019) and which, theoretically, should have the strongest link to harm-maximizing behavior in moral dilemmas about causing or preventing harm. In comparison to direct physical sadism, the relationship between both vicarious and verbal forms of sadism and whether one would engage in actions that directly cause or prevent physical harm to others is arguably more distal. To the extent that direct physical sadism is a better predictor of harm-maximizing tendencies than vicarious and verbal forms of sadism in scenarios involving decisions that directly cause or prevent harm, our finding that sadism is positively associated with sensitivity to consequences may simply reflect vicariously and verbally sadistic individuals' reduced aversion to otherwise difficult utilitarian decisions to sacrifice others. Research using more comprehensive measures of sadism is needed to unpack the associations of different facets of sadism with sensitivity to consequences.

6.3. Psychopathy and Moral-Dilemma Judgments

Although the link between psychopathy and utilitarian judgments is widely accepted in the field, our study suggests that psychopathy is unrelated to sensitivity to consequences, regardless of whether we partialed out the effects of the other aversive and affiliative traits. This result is consistent with the findings of Luke at al.'s confirmatory

analyses (2022), but inconsistent with the same study's exploratory analyses that modelled the CNI model parameters as latent variables, as well as other research that used the CNI model (e.g., Gawronski et al., 2017; Körner et al., 2020; Luke & Gawronski, 2021b). In addition, while our finding that psychopathy is negatively correlated with sensitivity to moral norms is consistent with past research (Gawronski et al., 2017; Körner et al., 2020; Luke & Gawronski, 2021b; Luke et al., 2022; Ng et al., 2022), the association disappeared after controlling for the other traits and after excluding potential outliers (see Supplemental Information).

As discussed by Luke and colleagues (2022), mixed associations between psychopathy and moral-dilemma judgments may be attributed to the different measures of psychopathy used across studies. Most prior research that used the CNI model (Gawronski et al., 2017; Körner et al., 2020; Luke & Gawronski, 2021b) assessed primary psychopathy using the Levenson's Self-Report Psychopathy Scale (Levenson et al., 1995). A notable exception is a study by Luke et al. (2022) which studied all four facets of psychopathy using the Self-Report Psychopathy Scale-Short Form. Compared to these measures, the SD4 is a relatively short measure of psychopathy that only captures the behavioral (e.g., *'I tend to fight against authorities and their rules.'*) and interpersonal (e.g., *'People who mess with me always regret it.'*) facets of the trait. Notably, the SD4 does not include any items assessing the affective deficits characteristic of psychopathy, which is a critical omission given that the affective facet of psychopathy has been found to uniquely predict sensitivity to moral norms (Luke et al., 2022). Whether our findings that psychopathy was not associated with both sensitivity to moral norms and sensitivity to consequences can be attributed to the fact that sadism is a stronger predictor or because we did not capture all facets of psychopathy remains unclear. Future research fully capturing the nuances of the aversive traits is needed to clarify the associations.

6.4. Implications for Affiliative 'Light' Traits Research

Our finding that the superordinate aversive and affiliative trait factors are both significant predictors of moral-dilemma judgments suggests that the traits are not mere opposites. Instead, they both uniquely influence our receptiveness to moral norms of harm. As discussed by Luke and Gawronski (2021b), that aversive traits negatively predict norm adherence may reflect either (1) a lack of concern for whether one's behavior is incongruent with societal expectations, (2) a deficient understanding of conventions involving moral norms, (3) or both. In the opposite vein, that affiliative traits positively predict norm adherence may reflect greater consideration of social judgments, with individuals with elevated affiliative traits being more likely to abide by a rule perceived to be in line with the majority view. Such an account would be consistent with past research demonstrating that individuals with elevated affiliative traits (and few aversive traits) show greater willingness to affiliate and trust others than their counterparts with other affiliative and aversive trait profiles (Neumann et al., 2020).

However, the current research raises important questions about what the traits captured by the LTS truly reflects. Our findings that (1) the subordinate affiliative traits did not independently predict any of the three CNI model parameters and (2) the superordinate affiliative trait was positively linked to sensitivity to norms suggest that norm adherence may have been driven by a general prosocial propensity. However, research by Lukić and Živanović (2021) has demonstrated that sadism is a strong predictor of the LTS. This result is consistent with our finding that, despite the significant associations of both the superordinate aversive and affiliative traits and moral-dilemma judgments, sadism emerged as the only subordinate trait that significantly predicted responses in moral dilemmas. This result is also consistent with our findings of the ESEM, which suggested potential cross-loadings of sadism and narcissism onto the latent affiliative factor. Does the Light Triad construct, as postulated by Lukić and Živanović, not only capture our prosocial tendencies to care for others, but also our aversion towards sadism then? Research jointly assessing the factor structure of the SD4 and LTS is needed to unpack what the LTS's Light Triad construct truly captures, particularly since the original study that identified the three affiliative traits did not include a measure of sadism (Kaufman et al., 2019).

In addition, our correlational analyses revealed positive associations between the aversive trait of narcissism and the two affiliative traits of humanism and faith in humanity, regardless of whether potential outliers were excluded from analyses. In the ESEM analyses, narcissism also weighed more heavily than Kantianism in estimating the latent affiliative factor. These statistically significant, positive associations between narcissism and affiliative traits run counter to the findings of Kaufman and colleagues (2019) whose research revealed the opposite pattern of results. One possible explanation for the unexpected positive associations is that participants with elevated narcissistic traits could have been especially concerned about social desirability and thus sought to present themselves in a more positive light, which is a behavioral tendency observed in prior research (Kowalski et al., 2018; Zuo et al., 2016). If so, research on affiliative traits would need to account for potential socially desirable responding.

6.5. Limitations

Before concluding, it is worth addressing some limitations of the current research. In terms of generalizability, the sample of this exploratory study was limited to participants from the United Kingdom and the sample predominantly identified as being White. Future research examining the relations between moral-dilemma judgments and aversive and affiliative traits could benefit from using more diverse samples. In terms of study measurement, although the SD4 allows us to measure the four subordinate aversive traits in a highly efficient manner, this convenience comes at a cost: as aforementioned, the SD4 subscale of sadism fails to capture direct physical sadism, which is likely to be a stronger predictor of antisocial tendencies than vicarious forms of sadism, and the SD4 subscale of psychopathy fails to encapsulate the important affective deficits that characterize the trait. The internal consistencies of the subscales are also not ideal, with the Kantianism subscale likely requiring further revision. Following this exploratory study, research using more nuanced and reliable measures of aversive and affiliative traits is needed.

7. Conclusion

Addressing two methodological limitations of extant work, the current research explored the associations of aversive and affiliative traits with moral-dilemma judgments. Our findings suggest that affiliative traits predict moral-dilemma judgments above and beyond aversive traits, but questions remain about what the affiliative Light Triad construct truly reflects. Our finding that sadism was the only subordinate trait significantly predicting sensitivity to consequences and sensitivity to moral norms highlights the gravity of overlooking sadism in aversive traits research (i.e., research using the Dark Triad of aversive traits). It further suggests that the extent to which we derive enjoyment from cruelty appears to play a central role in determining responses to moral scenarios involving serious physical harm. Future work investigating whether sadism would emerge as the strongest predictor of moral judgment and behavior in other types of moral scenarios (e.g., charitable donations) would also be meaningful in clarifying the extent to which our penchant for or aversion toward sadism shapes our morality.

CRediT authorship contribution statement

Nyx L. Ng: Conceptualization, Investigation, Formal analysis, Visualization, Writing – original draft. **Craig S. Neumann:** Conceptualization, Formal analysis, Funding acquisition, Writing – review & editing. **Dillon M. Luke:** Conceptualization, Writing – review & editing.

Bertram Gawronski: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The preregistration, data, analysis codes, and materials for the study can be accessed at https://osf.io/v3jxc/.

Appendix A. Supplemental Information

Supplemental information for this article can be found online at https://doi.org/10.1016/j.jrp.2023.104450.

References

Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2021). Themes of the dark core of personality. *Psychological Assessment, 33*(6), 511–525. https://doi.org/10.1037/pas0001006

Baron, J., & Goodwin, G. P. (2021). Consequences, norms, and inaction: Response to Gawronski et al. (2020). *Judgment and Decision Making, 16*(2), 31.

Baron, J., & Goodwin, G. P. (2020). Consequences, norms, and inaction: A critical analysis. *Judgment and Decision Making, 15*(3), 421–442.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*(1), 154–161. https://doi.org/10.1016/j.cognition.2011.05.010

Blais, J., Solodukhin, E., & Forth, A. E. (2014). A meta-analysis exploring the relationship between psychopathy and instrumental versus reactive violence. *Criminal Justice and Behavior, 41*(7), 797–821. https://doi.org/10.1177/0093854813519629

Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science, 24*(11), 2201–2209. https://doi.org/10.1177/0956797613490749

Campbell, J., Schermer, J. A., Villani, V. C., Nguyen, B., Vickers, L., & Vernon, P. A. (2009). A behavioral genetic study of the Dark Triad of personality and moral development. *Twin Research and Human Genetics, 12*(2), 132–136. https://doi.org/10.1375/twin.12.2.132

Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience, 5*(1), 59–67. https://doi.org/10.1093/scan/nsp051

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*(2), 216–235. https://doi.org/10.1037/a0031021

Crone, D. L., & Laham, S. M. (2017). Utilitarian preferences or action preferences? De-confounding action and moral code in sacrificial dilemmas. *Personality and Individual Differences, 104*, 476–481. https://doi.org/10.1016/j.paid.2016.09.022

Deutchman, P., & Sullivan, J. (2018). The Dark Triad and framing effects predict selfish behavior in a one-shot Prisoner's Dilemma. *PLOS ONE, 13*(9), Article e0203891. https://doi.org/10.1371/journal.pone.0203891

Dinić, B. M., Milosavljević, M., & Mandarić, J. (2021). Effects of Dark Tetrad traits on utilitarian moral judgement: The role of personal involvement and familiarity with the victim. *Asian Journal of Social Psychology, 24*(1), 48–58. https://doi.org/10.1111/ajsp.12422

Djeriouat, H., & Trémolière, B. (2014). The Dark Triad of personality and utilitarian moral judgment: The mediating role of Honesty/Humility and Harm/Care. *Personality and Individual Differences, 67*, 11–16. https://doi.org/10.1016/j.paid.2013.12.026

Doerfler, S. M., Tajmirriyahi, M., Dhaliwal, A., Bradetich, A. J., Ickes, W., & Levine, D. S. (2021). The Dark Triad trait of psychopathy and message framing predict risky decision-making during the COVID-19 pandemic. *International Journal of Psychology, 56*(4), 623–631. https://doi.org/10.1002/ijop.12766

Egan, V., Hughes, N., & Palmer, E. J. (2015). Moral disengagement, the Dark Triad, and unethical consumer attitudes. *Personality and Individual Differences, 76*, 123–128. https://doi.org/10.1016/j.paid.2014.11.054

Ellen, B. P., Alexander, K. C., Mackey, J. D., McAllister, C. P., & Carson, J. E. (2021). Portrait of a workplace deviant: A clearer picture of the Big Five and Dark Triad as predictors of workplace deviance. *Journal of Applied Psychology, 106*(12), 1950–1961. https://doi.org/10.1037/apl0000880

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review, 5*. https://doi.org/10.1093/0199252866.003.0002

Foulkes, L. (2019). Sadism: Review of an elusive construct. *Personality and Individual Differences, 151*, Article 109500. https://doi.org/10.1016/j.paid.2019.07.010

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass, 7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology, 113*(3), 343–376. https://doi.org/10.1037/pspa0000086

Gawronski, B., Conway, P., Hütter, M., Luke, D. M., Armstrong, J., & Friesdorf, R. (2020). On the validity of the CNI model of moral decision-making: Reply to Baron and Goodwin (2020). *Judgment and Decision Making, 15*(6), 1054–1072.

Glenn, A. L., Raine, A., Schug, R. A., Young, L., & Hauser, M. (2009). Increased DLPFC activity during moral decision-making in psychopathy. *Molecular Psychiatry, 14*(10), 909–911. https://doi.org/10.1038/mp.2009.76

Glenn, A. L., & Sellbom, M. (2015). Theoretical and empirical concerns regarding the Dark Triad as a construct. *Journal of Personality Disorders, 29*(3), 360–377. https://doi.org/10.1521/pedi_2014_28_162

Grežo, M., & Adamus, M. (2022). Light and Dark core of personality and the adherence to COVID-19 containment measures: The roles of motivation and trust in government. *Acta Psychologica, 223*, Article 103483. https://doi.org/10.1016/j.actpsy.2021.103483

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology, 4*(1), 217–246. https://doi.org/10.1146/annurev.clinpsy.3.022806.091452

Hennig, M., & Hütter, M. (2020). Revisiting the divide between deontology and utilitarianism in moral dilemma judgment: A multinomial modeling approach. *Social Cognition, 118*(1), 22–56. https://doi.org/10.1037/pspa0000173

Heym, N., Firth, J., Kibowski, F., Sumich, A., Egan, V., & Bloxsom, C. A. J. (2019). Empathy at the heart of darkness: Empathy deficits that bind the Dark Triad and those that mediate indirect relational aggression. *Frontiers in Psychiatry, 10*, 95. https://doi.org/10.3389/fpsyt.2019.00095

Hilbig, B. E., Thielmann, I., Klein, S. A., Moshagen, M., & Zettler, I. (2021). The dark core of personality and socially aversive psychopathology. *Journal of Personality, 89*(2), 216–227. https://doi.org/10.1111/jopy.12577

Horsten, L. K., Moshagen, M., & Hilbig, B. E. (2023). On the (number of) aversive traits it takes to approximate the common core of aversive personality. *Scientific Reports, 13*(1), 15021. https://doi.org/10.1038/s41598-023-42115-z

Johnson, L. K., Plouffe, R. A., & Saklofske, D. H. (2019). Subclinical sadism and the Dark Triad: Should there be a Dark Tetrad? *Journal of Individual Differences, 40*(3), 127–133. https://doi.org/10.1027/1614-0001/a000284

Jonason, P. K., & Buss, D. M. (2012). Avoiding entangling commitments: Tactics for implementing a short-term mating strategy. *Personality and Individual Differences, 52*(5), 606–610. https://doi.org/10.1016/j.paid.2011.12.015

Jonason, P. K., Luevano, V. X., & Adams, H. M. (2012). How the Dark Triad traits predict relationship choices. *Personality and Individual Differences, 53*(3), 180–184. https://doi.org/10.1016/j.paid.2012.03.007

Jonason, P. K., Lyons, M., Bethell, E. J., & Ross, R. (2013). Different routes to limited empathy in the sexes: Examining the links between the Dark Triad and empathy. *Personality and Individual Differences, 54*(5), 572–576. https://doi.org/10.1016/j.paid.2012.11.009

Jonason, P. K., Strosser, G. L., Kroll, C. H., Duineveld, J. J., & Baruffi, S. A. (2015). Valuing myself over others: The Dark Triad traits and moral and social values. *Personality and Individual Differences, 81*, 102–106. https://doi.org/10.1016/j.paid.2014.10.045

Kant, I. (1785). *Groundwork of the metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press.

Jones, D. N., & Figueredo, A. J. (2013). The core of darkness: Uncovering the heart of the Dark Triad. *European Journal of Personality, 27*(6), 521–531. https://doi.org/10.1002/per.1893

Karandikar, S., Kapoor, H., Fernandes, S., & Jonason, P. K. (2019). Predicting moral decision-making with dark personalities and moral values. *Personality and Individual Differences, 140*, 70–75. https://doi.org/10.1016/j.paid.2018.03.048

Kaufman, S. B., Yaden, D. B., Hyde, E., & Tsukayama, E. (2019). The Light vs. Dark Triad of personality: Contrasting two very different profiles of human nature. *Frontiers in Psychology, 10*, 467. https://doi.org/10.3389/fpsyg.2019.00467

Kay, C. S., & Arrow, H. (2022). Taking an elemental approach to the conceptualization and measurement of Machiavellianism, narcissism, and psychopathy. *Social and Personality Psychology Compass, 16*(4), Article e12662. https://doi.org/10.1111/spc3.12662

Kay, C. S., & Arrow, H. (2023). Eight misconceptions about the elemental approach and aversive personality trait research: A response to Andrews and colleagues (2023). *Social and Personality Psychology Compass*. , Article 12769. https://doi.org/10.1111/spc3.12769

Kay, C. S., & Saucier, G. (2020). Deviating from the social consensus: Relations among the Dark Triad, moral normativity, and general social normativity. *Personality and Individual Differences, 159*, Article 109889. https://doi.org/10.1016/j.paid.2020.109889

Körner, A., Deutsch, R., & Gawronski, B. (2020). Using the CNI model to investigate individual differences in moral dilemma judgments. *Personality and Social Psychology Bulletin, 46*(9), 1392–1407. https://doi.org/10.1177/0146167220907203

Kowalski, C. M., Rogoza, R., Vernon, P. A., & Schermer, J. A. (2018). The Dark Triad and the self-presentation variables of socially desirable responding and self-monitoring. *Personality and Individual Differences, 120*, 234–237. https://doi.org/10.1016/j.paid.2017.09.007

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*(1), 151–158. https://doi.org/10.1037/0022-3514.68.1.151

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liu, C., & Liao, J. (2021). CAN algorithm: An individual level approach to identify consequence and norm sensitivities and overall action/inaction preferences in moral decision-making. *Frontiers in Psychology, 11*, Article 547916. https://doi.org/10.3389/fpsyg.2020.547916

Luke, D. M., & Gawronski, B. (2021a). Political ideology and moral dilemma judgments: An analysis using the CNI model. *Personality and Social Psychology Bulletin, 47*(10), 1520–1531. https://doi.org/10.1177/0146167220987990

Luke, D. M., & Gawronski, B. (2021b). Psychopathy and moral dilemma judgments: A CNI model analysis of personal and perceived societal standards. *Social Cognition, 39*(1), 41–58. https://doi.org/10.1521/soco.2021.39.1.41

Luke, D. M., & Gawronski, B. (2022). Big five personality traits and moral-dilemma judgments: Two preregistered studies using the CNI model. *Journal of Research in Personality, 101*, Article 104297. https://doi.org/10.1016/j.jrp.2022.104297

Luke, D. M., Neumann, C. S., & Gawronski, B. (2022). Psychopathy and moral-dilemma judgment: An analysis using the four-factor model of psychopathy and the CNI model of moral decision-making. *Clinical Psychological Science, 10*(3), 553–569. https://doi.org/10.1177/21677026211043862

Lukić, P., & Živanović, M. (2021). Shedding light on the Light Triad: Further evidence on structural, construct, and predictive validity of the Light Triad. *Personality and Individual Differences, 178*, Article 110876. https://doi.org/10.1016/j.paid.2021.110876

Marshall, J., Watts, A. L., & Lilienfeld, S. O. (2018). Do psychopathic individuals possess a misaligned moral compass? A meta-analytic examination of psychopathy's relations with moral judgment. *Personality Disorders: Theory, Research, and Treatment, 9*(1), 40–50. https://doi.org/10.1037/per0000226

Međedović, J., & Petrović, B. (2015). The Dark Tetrad: Structural properties and location in the personality space. *Journal of Individual Differences, 36*(4), 228–236. https://doi.org/10.1027/1614-0001/a000179

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review, 125*(5), 656–688.

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: A meta-analysis and critical review of the literature on the Dark Triad (Narcissism, Machiavellianism, and Psychopathy). *Perspectives on Psychological Science, 12*(2), 183–204. https://doi.org/10.1177/1745691616666070

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.).

Neumann, C. S., & Hare, R. D. (2008). Psychopathic traits in a large community sample: Links to violence, alcohol use, and intelligence. *Journal of Consulting and Clinical Psychology, 76*(5), 893–899. https://doi.org/10.1037/0022-006X.76.5.893

Neumann, C. S., Jones, D. N., & Paulhus, D. L. (2022). Examining the Short Dark Tetrad (SD4) across models, correlates, and gender. *Assessment, 29*(4), 651–667. https://doi.org/10.1177/1073191120986624

Neumann, C. S., Kaufman, S. B., ten Brinke, L., Yaden, D. B., Hyde, E., & Tsykayama, E. (2020). Light and dark trait subtypes of human personality – A multi-study person-centered approach. *Personality and Individual Differences, 164*, Article 110121. https://doi.org/10.1016/j.paid.2020.110121

Ng, N. L., Luke, D. M., & Gawronski, B. (2022). Prediction of dishonest behavior using the CNI model of moral dilemma judgment: A registered report. *Comprehensive Results in Social Psychology, 1–28*. https://doi.org/10.1080/23743603.2022.2102473

O'Boyle, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *Journal of Applied Psychology, 97*(3), 557–579. https://doi.org/10.1037/a0025679

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Pajevic, M., Vukosavljevic-Gvozden, T., Stevanovic, N., & Neumann, C. S. (2018). The relationship between the Dark Tetrad and a two-dimensional view of empathy. *Personality and Individual Differences, 123*, 125–130. https://doi.org/10.1016/j.paid.2017.11.009

Patil, I. (2015). Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion. *Journal of Cognitive Psychology, 27*(3), 349–366. https://doi.org/10.1080/20445911.2015.1004334

Paulhus, D. L. (2014). Toward a taxonomy of dark personalities. *Current Directions in Psychological Science, 23*(6), 421–426. https://doi.org/10.1177/0963721414547737

Paulhus, D. L., Buckels, E. E., Trapnell, P. D., & Jones, D. N. (2021). Screening for dark personalities: The Short Dark Tetrad (SD4). *European Journal of Psychological Assessment, 37*(3), 208–222. https://doi.org/10.1027/1015-5759/a000602

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Plouffe, R. A., Smith, M. M., & Saklofske, D. H. (2019). A psychometric investigation of the Assessment of Sadistic Personality. *Personality and Individual Differences, 140*, 57–60. https://doi.org/10.1016/j.paid.2018.01.002

Sevi, B., & Doğruyol, B. (2020). Looking from the bright side: The Light Triad predicts Tinder use for love. *Journal of Social and Personal Relationships, 37*(7), 2136–2144. https://doi.org/10.1177/0265407520918942

Sevi, B., Urganci, B., & Sakman, E. (2020). Who cheats? An examination of light and dark personality traits as predictors of infidelity. *Personality and Individual Differences, 164*, Article 110126. https://doi.org/10.1016/j.paid.2020.110126

Szabó, E., & Jones, D. N. (2019). Gender differences moderate Machiavellianism and impulsivity: Implications for Dark Triad research. *Personality and Individual Differences, 141*, 160–165. https://doi.org/10.1016/j.paid.2019.01.008

Trémolière, B., & Djeriouat, H. (2016). The sadistic trait predicts minimization of intention and causal responsibility in moral judgment. *Cognition, 146*, 158–171. https://doi.org/10.1016/j.cognition.2015.09.014

Zamora, J., Ungson, N. D., & Seidman, G. (2022). The end justifies the me: Self-interest moderates the relationship between Dark Triad traits and utilitarian moral decisions. *Personality and Individual Differences, 184*, Article 111134. https://doi.org/10.1016/j.paid.2021.111134

Zuo, S., Wang, F., Xu, Y., Wang, F., & Zhao, X. (2016). The fragile but bright facet in the Dark Gem: Narcissism positively predicts personal morality when individual's self-esteem is at low level. *Personality and Individual Differences, 97*, 272–276. https://doi.org/10.1016/j.paid.2016.03.076