The association between personality traits and third-party moral judgment: A preregistered study

Flora Schwartz, Hakim Djeriouat, Bastien Trémolière

▶ **To cite this version:**

Flora Schwartz, Hakim Djeriouat, Bastien Trémolière. The association between personality traits and third-party moral judgment: A preregistered study. Acta Psychologica, 2021, 219, 10.1016/j.actpsy.2021.103392 . hal-03356032

HAL Id: hal-03356032
https://hal.science/hal-03356032

Submitted on 22 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0001691821001426
Manuscript_b7dc2b01b31f3b6cfccc9fb9d9fcf1c5

The association between personality traits and third-party moral judgment: a preregistered study

Flora Schwartz[1], Hakim Djeriouat[2], & Bastien Trémolière[1]

[1]UNIV. NIMES, APSY-V, F-30021 Nîmes Cedex 1, France.

[2]UNIV. LYON 2 69676 Bron Cedex, France.

Corresponding author: flora.schwartz@unimes.fr

© 2021 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

1. Introduction

Human societies are grounded on a set of conventional and moral rules that promote large-scale cooperation (Tomasello & Vaish, 2013). Peaceful interactions are also maintained by legal systems which aim to constrain rule-breaking conducts and prevent their occurrence as much as possible. Oftentimes, the responsibility of judging the agent of a moral transgression is deferred to independent third-party judges to minimize partiality. A range of situational and individual factors are likely to influence the mental representation of the events and the agents, thereby contributing to judgment severity of third-parties. Understanding how individuals make moral judgments lies at the intersection between Law, Philosophy and Psychology and investigations over the past years have started to uncover the psychological mechanisms involved in third-party moral judgment.

1.1. A two-process model of third-party moral judgment

At the time of assessing a moral transgression, people mainly process two components to reach their judgment: Intention (whether or not the agent had the intent to hurt someone) and the outcome (whether or not the victim was hurt as a result of the action taken). Most studies to date have used sacrificial moral dilemmas in which participants are asked whether it is acceptable to harm someone to save several others. Responses to these dilemmas are said to be “utilitarian” when people are prone to endorse a harmful action to favor the “greater good” and “anti-utilitarian” or “deontological” when people don’t find it acceptable to harm someone to benefit others. Research findings from such studies have contributed to a largely accepted theory that explains how people reach moral decisions. According to the dual-process model of moral judgment (Cushman, 2008; Greene et al., 2009), analyses of the intention to harm and the outcome for the victim are processed by distinct and complementary cognitive systems: the outcome-based and intent-based processes. The fast and automatic outcome-based process focuses on the consequences of a moral transgression for the victim. Conversely, the intent-

based process focuses on the examination of the agent's intentions (i.e., utilitarian motive for the killing, malevolent intention) and is slower and more costly.

Critically, the output of the two processes are sometimes opposed, triggering a cognitive conflict in the psychological processes at work (Greene et al., 2004). A typical case in which a conflict is present is during the judgment of accidental moral transgressions, when the agent unintentionally harms someone else, resulting in more or less severe consequences. In this case, the fast outcome-based process focuses on the harm inflicted to the victim and would increase the severity of the judgment, while the intent-based process focuses on the agent's innocent intentions and would decrease judgment severity. The presence of a cognitive conflict seems specific to accidental harm; this conflict is not triggered in the case of attempted harm, for which there is no salient negative outcome. Interestingly, individual differences are observed in the weight people attribute to the perpetrator's intent or the transgression's outcome and this might be associated with individual differences in solving this conflict. It may actually depend on a range of factors, such as susceptibility to implicit bias (Kang et al., 2012), political ideology (Clark & Wink, 2012) or thinking dispositions (Gunnell & Ceci, 2010; Patil et al., 2021). For example, individual political ideology predicts some differences in punishment severity, with conservatives allowing harsher fines (Clark & Wink, 2012). As another example, people with a more intuitive thinking style are for instance more likely to render a guilty verdict when analyzing case files for which the defendants appear less attractive in photographs than people more analytic (Gunnell & Ceci, 2010). Thinking style is also associated with individual differences in intent-based analysis following an accidental moral transgression (Patil et al., 2021).

Interestingly, individual differences in personality are also good candidates as moderators of the way people resolve such a cognitive conflict when judging the agent of accidental harm. To our knowledge, although the literature has shown some relationships between personality and moral judgment in the framework of moral utilitarianism, research has almost overlooked the effect of personality on third-party moral judgment.

1.2. Personality and third-party moral judgments

Endorsement of moral foundations, such as authority, loyalty, sanctity, fairness and care has been related to individual differences in personality traits (Graham et al., 2009). More specifically, differences in moral foundations as a function of political ideology is thought to be associated with individual differences in openness to experience, with lower openness associated with more endorsement of relatively conservative moral foundations (Graham et al., 2009). The present study focuses on the moral foundation of harm/care. With this respect, previous studies have tried to bridge the gap between personality traits and verdict preferences for defendants who harmed others (Devine et al., 2001; Shaffer et al., 1986; Shaffer & Case, 1982). Yet, these investigations are limited to legal decisions in court, and less is known about how personality traits predict the relative importance given to the specific components underlying the judgment of everyday moral transgressions, namely intention, and the severity of the outcome.

Among the handful of studies on the issue, previous work has for instance suggested that people with dark personalities are less concerned with the victim's harm when they make third-party moral judgments (Djeriouat & Trémolière, 2014; Trémolière & Djeriouat, 2016), and people with a more deliberative thinking style are more likely to make judgments decisions whose consequences favor the greater good, that is, benefitting the many by intentionally harming the few (Behnke et al., 2020). Interestingly, recent work using moral dilemmas has found that emotionality predicted sensitivity to the consequence of an action, while honesty-humility predicted sensitivity to moral norms (Kroneisen & Heck, 2020). However, the joint contribution of the personality variables to different types of moral transgressions is unknown.

In the present research we targeted three specific personality traits, namely, honesty-humility, emotionality and conscientiousness which appear good candidates for the modulation

of third-party moral judgment. We expect that these traits may weigh in the balance between the outcome-based and intent-based processes and thus generate variability in judgment severity.

1.3. Emotionality and the outcome-based process

Emotionality is a personality trait directly associated with aversion to harm (Cushman & Young, 2009; Greene et al., 2009). Emotionality is related to the construct of kin altruism and comprises the notion of empathic concerns, emotional connection, harm-avoidance and help-seeking behaviors (Ashton & Lee, 2007). People high in emotionality experience more anxiety in response to dangers and life's challenges, and need emotional support from others. People with low emotionality are not worried in challenging situations, including those involving physical harm, and feel emotionally detached from others.

Because emotionality is directly associated with aversion to inflicting harm, this personality trait may specifically impact the outcome-based process of moral judgment. Emotionality has recently been related to the sensitivity for consequences in moral dilemmas (Kroneisen & Heck, 2020) and its role in the outcome-based process of moral judgment is further suggested by research highlighting the relationship between emotion processing and the judgment of moral dilemmas (Greene et al., 2001; Moll et al., 2003; Moll & de Oliveira-Souza, 2007), in which the victim's harm is likely to trigger an emotional reaction (Cushman & Young, 2009; Greene et al., 2001).

However, there has been limited research about how individual differences in emotionality shape responses to moral dilemmas. Evidence from subclinical populations suggests that emotional callousness is associated with less aversion to the perspective of the sacrificial option. For instance, higher utilitarianism has been reported in individuals with alexithymia (Patil & Silani, 2014a), a subclinical condition defined by low emotional interoception and low empathy (Lane et al., 2015) and often found in psychopaths and individuals with autism spectrum disorder (Bird & Cook, 2013). Consistent with this observation, subclinical

psychopathy seems to be related to higher acceptability of harm in moral dilemmas (Behnke et al., 2020; Karandikar et al., 2019).

Beyond the framework of moral dilemmas, recent investigations in populations low in emotionality also suggest a more complex interaction between emotionality and different types of harm. For example, subclinical psychopathy is associated with less severe punishment of an agent who harmed intentionally (Hart et al., 2020; Trémolière & Djeriouat, 2016) or attempted to harm (Djeriouat & Trémolière, 2014). Additionally, alexithymia was associated with higher acceptability of accidental harm in individuals with autism (Patil & Silani, 2014b). Finally, patients with ventro-medial prefrontal cortex (vmPFC) lesions, a brain region involved in emotion regulation, judged harmful intents less severely (Young et al., 2010). To sum up, these studies suggest that emotional callousness is related to variability in moral judgment. Specifically, low emotionality may be associated with less concern for the victim's harm overall and less consideration of the agent's motives, while high emotionality may be associated with higher concern for the victim and harsher punishment of the agent who harmed intentionally. Building upon research suggesting that moral judgment may be the final output of two competing processes (Cushman, 2008), one intuitive and another deliberative, we now turn to a personality trait which has been associated in the literature with deliberative thinking, that is, conscientiousness.

1.4. Conscientiousness and the intent-based process

People high in conscientiousness tend to be organized, disciplined, favor accuracy in their tasks, and carefully deliberate before making decisions (Ashton & Lee, 2009). Conversely, people low in conscientiousness tend to avoid difficult tasks and make decisions with little reflection. In the context of moral judgment, a high conscientiousness trait may be associated with careful examination of the victim's harm on the one hand, and of the agent's intention on the other hand, whereas low conscientiousness may be associated with fast and intuitive

decisions based on the victim's harm. This idea is suggested by behavioral and neuroimaging studies using moral dilemmas.

It has indeed been found that experimental manipulations that prevent deliberative thinking, such as time pressure (Suter & Hertwig, 2011) or cognitive load (Green et al., 2008) affect utilitarian decisions in moral dilemmas. Conversely, facilitating deliberative thinking may favor utilitarianism (Nichols & Mallon, 2006), and a more deliberative thinking style is accompanied by greater cost-benefit analysis when evaluating moral dilemmas (Bartels, 2008). Additionally, evidence from functional neuroimaging points to a greater involvement of cortical brain regions involved in working memory and cognitive control in participants who make more utilitarian responses to moral dilemmas (Greene et al., 2004). Taken together, these findings suggest the involvement of cognitive resources while making a moral judgment.

Because conscientiousness seems to play a role in mental shifting (Fleming et al., 2016) and is related to impulse control skills (Roberts et al., 2014), people scoring high on this trait may engage in a more thoughtful examination of intent before making a moral judgment. Additionally, studies on clinical and developmental populations also suggest that executive functioning has a great share in the ability to primarily examine intentional states of an agent and to disentangle mental state analysis from harm-related affect to produce moral judgements. For example, the development of executive functions in children accompanies the shift from outcome-based to intent-based moral judgment (Zelazo et al., 1996). Additionally, among individuals on the autism spectrum, who show mental shifting deficits and tend to be low in conscientiousness (Schriber et al., 2014), high-functioning autism was associated with the exoneration of agents who only attempted to harm, and with greater severity toward agents who accidentally harmed someone (Moran et al., 2011).

In light of the above evidence and the two-process model of moral judgment, high conscientiousness people may thus be able to shift more easily from the outcome-based to the

intent-based process and to examine the agent's mental state, especially when the intent and outcome go in opposite directions, as for accidental harm. But how conscientiousness influences punishment decisions and acceptability of different types of moral transgressions is still unknown.

1.5. Honesty/Humility and the interaction between intent and outcome

Because honesty-humility encompasses several facets of moral behavior, it is likely to play a role in third-party moral judgment. This trait subsumes various facets such as sincerity, modesty, fairness, and greed-avoidance (Ashton & Lee, 2009). Honesty-humility predicts the active aspect of prosocial behavior (such as help and cooperation; (Hilbig et al., 2013)). It represents fairness, sincerity and honesty versus pretentiousness, conceitedness and narcissism (Lee & Ashton, 2008). People who score low on honesty–humility feel entitled and motivated to attain profit by exploiting others (Exline & Geyer, 2004; Perugini et al., 2003).

There is evidence that people high on this trait may be harsher toward an agent who broke -or attempted to break- moral rules. Indeed, prosocial behavior may be accompanied with an increased need for justice and demand of higher compensation for victims of moral transgressions (Fehr & Gächter, 2002). Conversely, people with antagonistic personality traits, such as narcissism and Machiavellianism (who score low on honesty-humility), show atypical moral judgment for different types of transgressions. For example, people scoring high on the dark triad (Paulhus & Williams, 2002), or dark tetrad of personality (Furnham et al., 2013) are less severe toward an agent who harmed intentionally (Hart et al., 2020; Trémolière & Djeriouat, 2016) or who attempted to harm someone (Djeriouat & Trémolière, 2014). Besides, people under the subclinical range of sadism were harsher toward the agent of accidental harm (Trémolière & Djeriouat, 2016). In keeping with the idea that honesty-humility may play a role in aversion to harm (Djeriouat & Trémolière, 2014), this personality trait may predict harsher judgment when the intent to harm is present.

1.6. The present study

In the present pre-registered study, we aimed to determine the association between personality traits and third-party moral judgment, with a specific focus on honesty-humility, emotionality, and conscientiousness. We expected emotionality to predict the processing of the outcome, while we expected conscientiousness and honesty-humility to predict the processing of the agent's intent. We predicted that these personality traits will be most important in explaining inter-individual differences when the intent to harm and the outcome are opposed, especially for accidental harm. Specifically, we expected emotionality to positively correlate with severity of punishment and blame of the agent following accidental moral transgressions, and to negatively correlate with acceptability of the agent's behavior. Conversely, we expected conscientiousness to negatively correlate with severity of punishment and blame following accidental moral transgressions, and to positively correlate with acceptability of the agent's behavior. Finally, we expected honesty-humility to positively correlate with severity of punishment and blame following intentional and attempted moral transgressions, and to negatively correlate with acceptability of the agent's behavior.

1. Method

2.1. Research Strategy

Our research plan including the protocol and the hypotheses was pre-registered on the Open Science Framework platform prior to collecting the data. The data, scripts and pre-registration report are publicly available at the link appearing at the end of the manuscript.

2.2. Sample size rationale

We used the G*power software to estimate the minimum sample size for this study. We ran an a-priori power analysis for GLM (fixed model, R^2, deviation from 0) with 5 predictors, with alpha = 0.05, Power = 0.95, effect size f^2 set to 0.05 (small effect size expected). The required sample size to observe such effect size under these assumptions was 402 participants.

2.3. Participants

Four hundred and five participants aged 18 or older living in the United Kingdom were invited to complete an online survey on the Prolific service and were paid £7/hour upon completion. Participants who failed attention checks and/ or who were flagged as fraudulent respondents were excluded from analysis. Based on these criteria, we excluded 8 participants. The final sample included 397 participants (mean age = 33.9 +-13.4 years, 248 females).

2.4. Material

Moral judgment task. Eight short stories featuring moral transgressions borrowed from Young and colleagues (Young et al., 2010) were used. The stories framed an interaction between an agent and a victim in a daily-life context. The type of moral transgression was manipulated using a 2 x 2 within-subject design with intent (present/absent) and outcome (harm/no harm) as categorical factors. This led to 4 conditions: intentional harm (intent present, harmful outcome), attempted harm (intent present, no harm), accidental harm (no intent, harmful outcome) and neutral as a control condition (no intent, no harm). See Table 1 for an example of a framing that could lead to 4 possible items. (Full description of the scenarios and conditions of moral transgressions is displayed in Appendix A). For each scenario, participants were asked to judge the agent's behavior on three outcome measures, using a 7-point Visual analog scale: (1) how acceptable the agent's behavior is (from "not at all acceptable" to "very acceptable"), (2) how

much blame the agent deserves (from “not at all” to “very much”), and (3) how much punishment the agent deserves (from “not at all” to “very much”).

Background	Erica is driving home when she sees a runner by the side of the road. The runner is bent over and has one hand on his chest. There is an empty asthma inhaler on the ground.	
Foreshadow	Negative The runner is having a serious asthma attack and needs to get to a hospital immediately before he goes into shock.	Neutral The runner is just short of breath and has stopped to rest for a moment before continuing his long jog.
Intent	Negative Because Erica sees the inhaler and the gasping runner, she believes that the runner is having an asthma attack and must get to a hospital immediately.	Neutral Because Erica doesn't see the inhaler, she believes that the runner has just stopped to catch his breath and will continue his jog shortly.
Outcome	Negative Erica continues to drive, leaving the runner. The runner faints and collapses.	Neutral Erica continues to drive, leaving the runner. The runner catches his breath and is fine.

Table 1. Item versions in the moral judgment task. Notes: The same framing leads to 4 intent by outcome combinations. A negative (respectively, neutral) foreshadow always precedes a negative (respectively, neutral) outcome.

Personality traits. Personality traits were assessed using two self-report questionnaires: the HEXACO personality inventory and the sensation seeking scale 5. The HEXACO-60 inventory (Ashton & Lee, 2009) is a 60-item questionnaire that assesses 6 major personality traits (each consisting of 10 items): Honesty/Humility (e.g., “I would not use flattery to get a rise or a

promotion at work, even if I thought it would succeed"), Emotionality (e.g., "I would feel afraid if I had to travel in bad weather conditions"), Extraversion (e.g., "I rarely express my opinion in group meetings"), Agreeableness (e.g., "People sometimes tell me that I am too critical of others"), Conscientiousness (e.g., "I plan ahead and organize things, to avoid scrambling at the last minute"), and Openness to experience (e.g "I would enjoy creating a work of art, such as a novel, a song or a painting"). Participants rated their agreement with each item using a 5-point Likert scale, ranging from "Strongly disagree" to "Strongly agree". Means scores are determined for each of the 6 dimensions. All personality dimensions showed an internal consistency (Cronbach's alpha) above 0.77.

2.5. Procedure

Participants were told they would be asked questions about the behavior of one of the story's characters. They were presented with the moral judgment task immediately before or after completing the self-report questionnaires. Task order was counterbalanced across participants and had no impact on the main results. Participants were presented with 8 moral transgression items (2 items / condition). The presentation of items was pseudo-randomized. First, the association between framing and type of moral transgression was counterbalanced across participants. Second, the order of items was fully randomized. After completion of the moral judgment task, participants responded to attention checks asking whether they responded to the best of their ability (Yes / no / maybe / I don't want to answer) and whether they answered without reading the story (Yes / no / maybe / I don't want to answer).

2.6. Data analysis

Responses to the moral judgment task (level of acceptability, blame, and punishment) were analyzed using a within-subject design with "intent" and "outcome" as within-subject factors. We

used continuous predictors (scores at the HEXACO) to determine how individual differences in moral judgment are explained by personality traits for the different intent by outcome combinations. One item from the moral judgment task ("seatbelt") was excluded from the analysis because the control condition did not show the expected response pattern (low acceptability overall for the neutral condition, while acceptability of the neutral condition was homogeneously high for all other items).

We implemented linear mixed models following the steps outlined by Zuur (Zuur et al., 2009). We first assessed the relevance of using mixed models by computing a general linear model with the factors described above. As the residual variance was large enough, we constructed a series of mixed models. We determined the optimal random effects structure and then the optimal fixed structure. Since models with random slopes failed to converge, we used only random intercepts. Moreover, based on the AIC criteria, the different types of moral transgressions were coded as one 4-level factor (instead of being defined as the "intent" and "outcome" factors). The final model included type of transgression, honesty-humility, emotionality and conscientiousness as fixed factors, with by-subject and by-item random intercepts. The neutral condition (=control) was coded as the reference level of the type of transgression factor. Interaction terms between type of transgression and each of the three personality traits were included in the final model. The final model reads:

```
moral evaluation ~ transgression + honesty + emotionality + conscientiousness
+ transgression * honesty + transgression * emotionality + transgression *
conscientiousness + (1|item) + (1|subject)
```

Statistical significance of fixed effects and their interaction was obtained using the Satterthwaite approximation implemented in the lmerTest R package (Kuznetsova et al., 2017). The Satterthwaite approximation is one of the methods recommended for studies including a limited number of items (Luke, 2017), as in our case. Post-hoc pairwise comparisons were performed using Tukey tests from the lsmeans package (Lenth, 2016). We also ran exploratory

analyses to examine moral evaluation as a function of all 6 personality traits from the HEXACO as well as sensation seeking questionnaires.

1. Results

3.1. Manipulation check

Participants' moral judgment for the different types of transgressions was in accordance with the literature on moral psychology. Across evaluations of acceptability, punishment and blame, intentional transgressions were as expected the most severely judged, followed by attempted, and then accidental transgressions, while the most lenient judgment was consistently made for the neutral condition (See Supplementary results and Supplementary Figure 1 for detailed information).

3.2. Individual personality traits and moral judgment

To better visualize how moral judgment is predicted by specific personality traits independently of one another, we first ran Pearson's correlations between moral judgment (acceptability, punishment and blame) associated with each type of transgression (accidental, attempted, intentional, neutral), and all personality measures. Results are presented in Appendix B.

Next, in keeping with our preregistration, we determined the association between 3 of these personality traits and moral judgment in our main analyses (emotionality, conscientiousness, and honesty-humility). We first implemented linear mixed models for each type of moral evaluation (acceptability, punishment and blame), with type of transgression as categorical factor, personality traits (honesty-humility, emotionality, and conscientiousness) as continuous predictors and by-item and by-subject random intercepts. We report the main effects and interaction terms of the full model. Secondly, we implemented linear mixed models for each

type of transgression separately, with personality traits as fixed factors and by-item and by-subject random intercepts. This step allowed us to assess more specifically how personality traits relate to moral judgment for each type of transgression.

First, emotionality predicted acceptability ratings overall ($F = 3.83$, $p = 0.051$). More specifically, higher emotionality was associated with lower acceptability of attempted transgressions ($b = -0.029$, $SE = 0.010$, $p = 0.0043$) (Figure 1, top panel). Moreover, although the main effect of emotionality did not reach significance for decisions of blame ($F = 2.09$, $p = 0.15$) and punishment ($F = 1.28$, $p = 0.26$), the interaction with the type of transgression was significant for blame ($F = 3.10$, $p = 0.026$) and marginally significant for punishment ($F = 2.18$, $p = 0.088$). In the case of intentional transgressions, higher emotionality marginally predicted an increase in blame ($b = 0.017$, $SE = 0.009$, $p = 0.074$) and punishment ($b = 0.019$, $SE = 0.01$, $p = 0.053$) (Figure 1, central and bottom panels).

Second, conscientiousness did not significantly predict overall ratings of acceptability ($F = 0.35$, $p = 0.56$), punishment ($F = 0.060$, $p = 0.81$) and blame ($F = 0.039$, $p = 0.84$). However, higher conscientiousness was associated with lower acceptability of the agent's behavior following intentional transgressions specifically ($b = -0.017$, $SE = 0.0077$, $p = 0.027$) (Figure 1, top panel).

Third, we found a main effect of honesty-humility on acceptability of the agent's behavior ($F = 15.67$, $p < 0.001$). Honesty-humility negatively predicted acceptability of accidental ($b = -0.030$, $SE = 0.012$, $p = 0.017$), and neutral transgressions ($b = -0.022$, $SE = 0.0091$, $p = 0.016$), and tended to do so for attempted transgressions ($b = -0.017$, $SE = 0.010$, $p = 0.099$), but did not significantly predict acceptability of intentional transgressions ($b = -2.10^{-5}$, $SE = 0.0068$, $p = 0.99$) (Figure 1, top panel). However, the main effect of honesty-humility was not significant on blame ($F = 2.02$, $p = 0.15$) and punishment ($F = 1.21$, $p = 0.27$), and this trait did relate to punishment and blame judgments for any transgression (Figure 1, central and bottom panel).

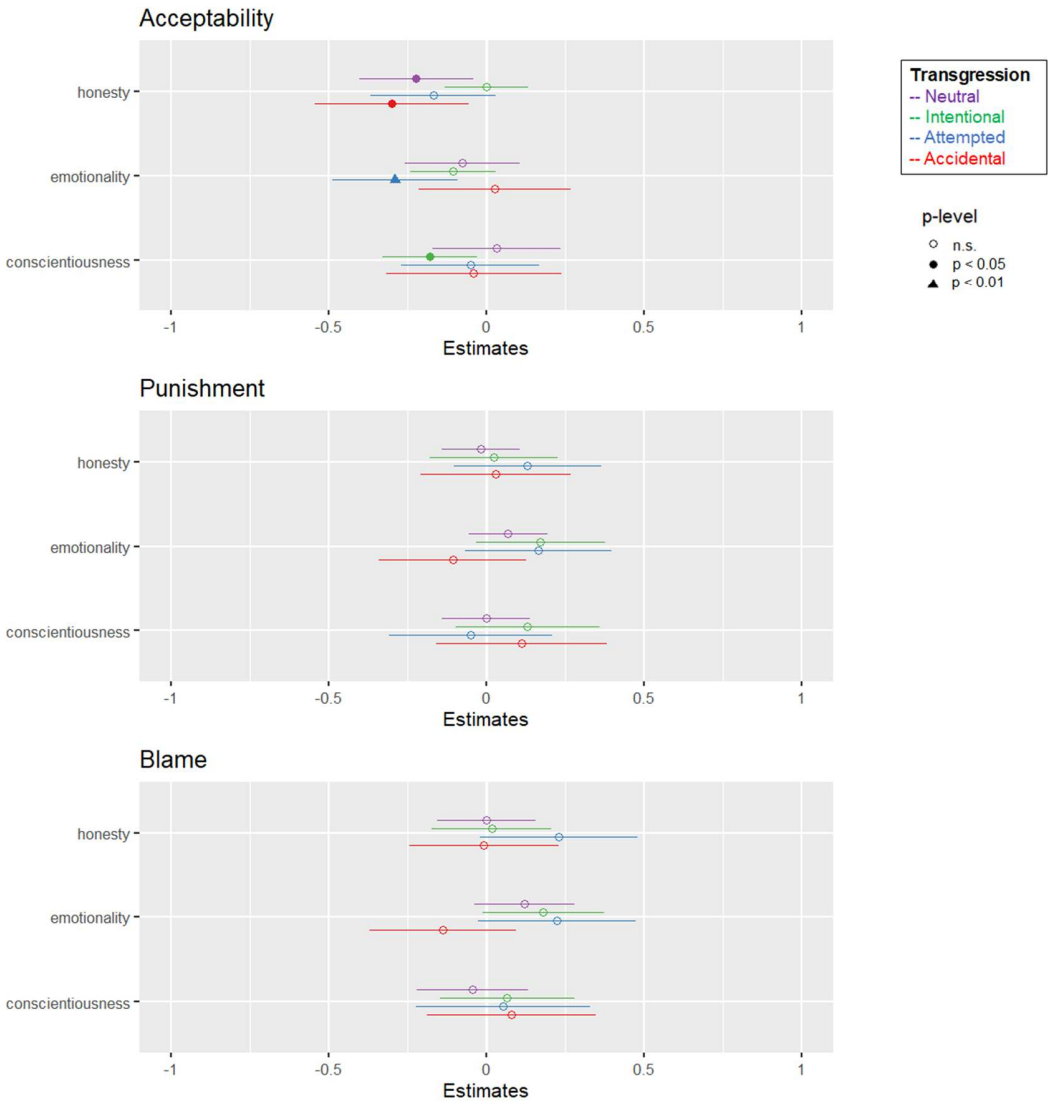

Figure 1. Prediction of moral judgments by Honesty-Humility, Emotionality and Conscientiousness. Horizontal bars represent the 95% Confidence Interval of the mean estimate.

1. Exploratory analyses

In the pre-registered exploratory analyses, we added the remaining personality variables from the HEXACO (openness to experience, agreeableness, extraversion) in the mixed models

in order to account for a larger range of personality traits. After running a posteriori sensitivity analyses, we found that the results reported below were large enough to be detected with sufficient statistical power. Importantly, in addition to preserving the specific link between honesty-humility, emotionality and conscientiousness and judgments of acceptability (as reported previously), inclusion of the remaining personality traits revealed (1) the significant relationship between extraversion and judgments of blame ($F = 7.87$, $p = 0.005$) and punishment ($F = 4.01$, $p = 0.046$), (2) the interaction between openness to experience and type of transgression for acceptability ratings ($F = 3.53$, $p = 0.014$), explained by a negative correlation between openness to experience and acceptability of the agent's behavior following attempted transgressions ($b = -0.024$, $SE = 0.012$, $p = 0.038$), and (3) enhanced the association between emotionality and overall acceptability ($F = 4.95$, $p = 0.027$) as well as decisions of blame ($b = 0.020$, $SE = 0.010$, $p = 0.053$) and punishment ($b = 0.019$, $SE = 0.011$, $p = 0.072$) following intentional transgressions specifically (Figure 2). In exploratory analyses, we determined whether sensation seeking predicted moral judgment beyond the hexaco traits and found no evidence thereof (see Supplementary results). In unplanned exploratory analyses, we determined whether the influence of personality traits on moral judgment was dependent on gender. As gender differences have been reported in honesty-humility and emotionality (Weller et al., 2018), we controlled for this potential confound (supplementary results). The main effect of gender was not significant, but gender interacted with conscientiousness for judgments of acceptability ($F = 4.34$, $p = 0.038$).

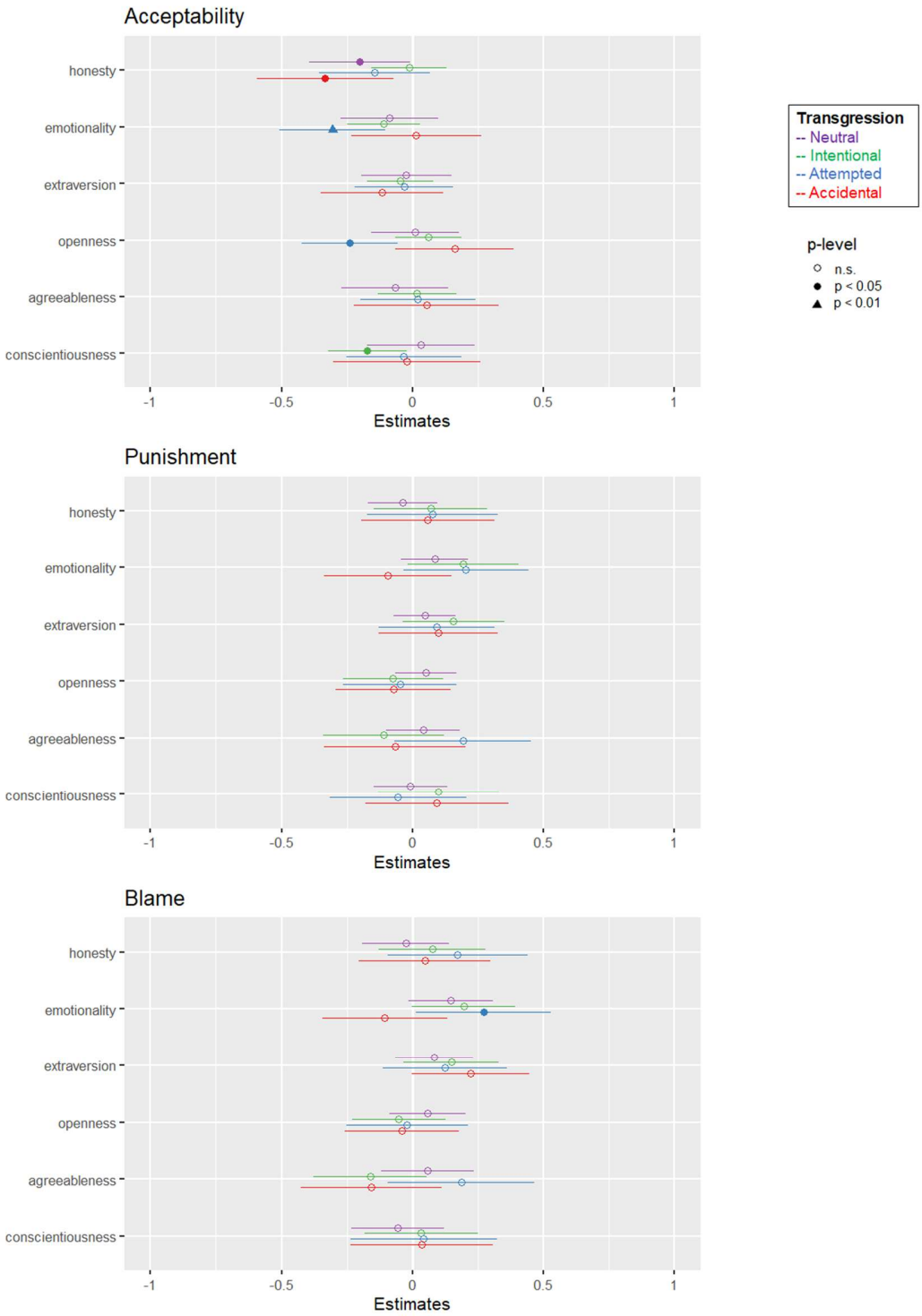

Figure 2. Prediction of moral judgments by the HEXACO personality traits. Horizontal bars represent the 95% Confidence Interval of the mean estimate.

4. Discussion

We investigated the relationship between personality traits and third-party moral judgment. Across different intent by harm combinations, we found that judgment severity for the different types of transgressions was selectively predicted by personality variables. While higher honesty-humility scores were associated with lower acceptability of moral transgressions overall, emotionality and openness to experience were negatively correlated with acceptability ratings when the agent attempted to harm, and conscientiousness was negatively correlated with acceptability of intentional harm. We also found that emotionality was moderately associated with increased punishment and blame of an agent who harmed intentionally.

4.1. Emotionality modulates the intent-based process of moral judgment

Our finding that emotionality predicts moral judgment is in line with the literature that used moral dilemmas in typical populations (Conway et al., 2018; Crockett et al., 2010; Greene et al., 2004; Suter & Hertwig, 2011; Kroneisen & Heck, 2020) as well as with evidence of atypical moral judgment in subclinical or clinical populations with emotional deficits (Djeriouat & Trémolière, 2014; Karandikar et al., 2019; Patil & Silani, 2014b; Young et al., 2010). Alexithymia has for instance been associated with less concern for the victim (Patil & Silani, 2014a) and psychopathy has been associated with less concern over a victim's death (Behnke et al., 2020).

However, contrary to our hypothesis, we found that emotionality predicts acceptability of attempted harm, but not accidental harm. Moreover, after adding extraversion to the model in exploratory analyses, emotionality also tended to predict punishment decisions and blame of intentional harm. These results suggest that emotionality may sometimes modulate the intent-

based process as much as (or even more so than) the outcome-based process when deciding about moral wrongness. This is consistent with some recent revisions of the dual-process model of moral judgment (Cushman, 2013), which posit that affect may influence both processes. The present findings further converge with the recent observation that distinct emotions may be triggered by the intent to harm on the one hand, and the victim's harm on the other hand (Hechler & Kessler, 2018). More specifically, anger at an agent who intends to harm is distinct from empathic concern for the victim (Hechler & Kessler, 2018).

Why did emotionality fail to predict judgment of accidental transgressions in the current study? We see two potential explanations: the contact principle and the action/ omission distinction, which are critical to the judgment of moral transgressions (Cushman, 2013). First, the contact principle of moral judgment suggests that someone who inflicted harm directly (by physical force) is judged more severely than someone who harmed indirectly (Cushman, 2013; Greene et al., 2009). In short, not only the presence of harm induces an emotional response, but also how it has been done (i.e., giving someone poisonous food versus beating someone). The fact that harm is not inflicted directly in all our scenarios may explain why individual differences in emotionality don't explain the judgment severity of accidental harm. Second, when (accidental) harm results from an action, the transgression is judged more severely than when (accidental) harm results from an omission, an effect known as the "omission bias" (Baron & Ritov, 2004; Spranca et al., 1991). In our study, the victim's harm resulted mostly from omissions. As a consequence, emotionality may have predicted moral condemnation to a lesser extent for harmful omissions than it would have for harmful actions.

4.2. Conscientiousness is associated with more severe judgment of moral wrongness following intentional harm

People high in conscientiousness favor deliberative thinking more than others when making a decision (Ashton & Lee, 2009), and we expected this to also be the case for moral judgment. More specifically, we reasoned that the careful examination of intentionality may play a greater role in people with high conscientiousness. Accordingly, we expected conscientiousness to predict the judgment of accidental harm especially, because this is a situation in which the output of the intent-based and outcome-based processes are opposed and whereby deliberation is presumably needed to solve the conflict (Greene, 2007; Greene et al., 2004). However, conscientiousness did not relate to judgment of accidents in the present study. Rather, we found that conscientiousness was associated with less acceptability of the agent's behavior following intentional transgressions.

Previous studies on moral dilemmas have found that greater deliberation was associated with greater acceptability of inflicting harm for utilitarian motives (Bartels, 2008; Behnke et al., 2020), and accepting harm as a mean to preserve the greater good was cognitively demanding (Greene et al., 2008; Suter & Hertwig, 2011; Trémolière et al., 2012). The present study complements this literature by suggesting that deliberative thinking may additionally modulate moral outrage following unfair or evitable harm, that is, when an agent harmed intentionally without utilitarian motives.

4.3. Honesty-Humility modulates moral outrage, but not punishment decisions

We expected honesty-humility to be more sensitive to the intent to harm, and thus to be associated with more severe judgment of intentional and attempted transgressions. On the contrary, honesty-humility was negatively correlated with acceptability of the agent's behavior following accidental transgressions.

This finding seems hardly reconcilable with previous evidence indicating that people are particularly sensitive to the unfairness component of harmful events. Moral outrage following

harm is, for instance, higher if harm serves unfair motives (Fehr & Gächter, 2002), and prosocial attitude is associated with a greater need for justice in third parties in order to compensate for unfair harm (Pfattheicher et al., 2019). Honesty-humility also predicts desirability of revenge (Edwin Sheppard & Boon, 2012). Despite this indirect evidence that honesty-humility may be associated with harsher judgment of intentional and attempted transgressions, individual variability in this trait failed to predict outrage and condemnation of a malevolent agent in the current study.

By contrast, this finding may support the idea that honesty-humility is related to aversion to actual harmful outcomes rather than aversion to harmful intents. Such aversion may thus predict greater moral outrage when an agent accidentally harms another. In people with dark personalities, who are less reluctant to inflict harm, honesty-humility indeed seems to reduce utilitarianism (Djeriouat & Trémolière, 2014). However, since honesty-humility also negatively predicted acceptability in the neutral condition, this result should be interpreted with caution.

Interestingly, honesty-humility predicted judgments of acceptability, but not judgments of punishment and blame. This suggests that individual variability in this trait is more predictive of moral outrage than it is of punishment. Even though punishment decisions and moral outrage at third parties are correlated overall (Hechler & Kessler, 2018), individuals with high honesty-humility may not necessarily punish an agent in proportion to the harm done or to the outrage triggered. Speculatively, because honesty-humility is associated with cooperation (Hilbig et al., 2013), concern for justice in individuals scoring high on this trait may be more oriented toward compensation for the victim and rehabilitation of the agent, and less so toward retributive punishment.

5. Conclusion

In sum, we found that personality may play a moderate, yet selective role in third-party moral judgment. Our study suggests that some personality traits may contribute both to the intent-based and outcome-based processes of moral judgment, while others may modulate one of these two processes. With this respect, it calls for further investigation of the sub-processes underlying the examination of intent and outcome. Future work should study the interaction between individual dispositions and contextual factors in third-party moral judgment. An interesting avenue is for instance how identification with the agent or the victim modulates judgment of different types of transgressions.

Conflict of interest

The authors declare no conflict of interest

Informed consent.

Participants completed the experiment only if they wished to do so.

All procedures performed in this study were in accordance with the 1964 Helsinki declaration and its later amendments. This study respected the ethics code of the American Psychological Association (2002).

Acknowledgments

This work was supported by an ANR grant (ANR-19-CE28-0002).

Data availability statement

The data, scripts and pre-registration are available at the following webpage: https://osf.io/n6ht3

References

Ashton, M. C., Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*, 150-166.

Ashton, M., & Lee, K. (2009). The HEXACO-60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Babalola, M. T., Bligh, M. C., Ogunfowora, B., Guo, L., & Garba, O. A. (2019). The Mind is Willing, but the Situation Constrains: Why and When Leader Conscientiousness Relates to Ethical Leadership. *Journal of Business Ethics*, *155*(1), 75–89. https://doi.org/10.1007/s10551-017-3524-4

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*(2), 74–85. https://doi.org/10.1016/j.obhdp.2004.03.003

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*(2), 381–417. https://doi.org/10.1016/j.cognition.2008.03.001

Behnke, A., Strobel, A., & Armbruster, D. (2020). When the killing has been done: Exploring associations of personality with third-party judgment and punishment of homicides in moral dilemma scenarios. *PLOS ONE*, *15*(6), e0235253. https://doi.org/10.1371/journal.pone.0235253

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, *3*, e285. PubMed. https://doi.org/10.1038/tp.2013.61

Brady, J. P. (1983). Fair and impartial railroad: The jury, the media, and political trials. *Journal of Criminal Justice*, *11*(3), 241–263. https://doi.org/10.1016/0047-2352(83)90117-4

Clark, J. W., & Wink, K. (2012). The relationship between political ideology and punishment: What do jury panel members say? *Applied Psychology in Criminal Justice*, *8*(2), 130–146.

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process

dissociation and the judgments of philosophers. *Cognition*, *179*, 241–265. https://doi.org/10.1016/j.cognition.2018.04.018

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*(40), 17433–17438. https://doi.org/10.1073/pnas.1009396107

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., & Young, L. (2009). The Psychology of Dilemmas and the Philosophy of Morality. *Ethical Theory and Moral Practice*, *12*(1), 9–24. https://doi.org/10.1007/s10677-008-9145-3

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law*, *7*(3), 622–727. https://doi.org/10.1037/1076-8971.7.3.622

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

Edwin Sheppard, K., & Boon, S. D. (2012). Predicting appraisals of romantic revenge: The roles of Honesty–Humility, Agreeableness, and vengefulness. *Personality and Individual Differences*, *52*(2), 128–132. https://doi.org/10.1016/j.paid.2011.09.014

Exline, J. J., & Geyer, A. L. (2004). Perceptions of Humility: A Preliminary Study. *Self and Identity*, *3*(2), 95–114. https://doi.org/10.1080/13576500342000077

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. https://doi.org/10.1038/415137a

Fleming, K. A., Heintzelman, S. J., & Bartholow, B. D. (2016). Specifying Associations Between Conscientiousness and Executive Functioning: Mental Set Shifting, Not Prepotent Response Inhibition or Working Memory Updating: Specifying Associations

Between Conscientiousness and Executive Functioning. *Journal of Personality*, *84*(3), 348–360. https://doi.org/10.1111/jopy.12163

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

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, *96*(5), 1029.

Green, M., Piel, J. A., & Flowers, C. (2008). Reversing Education Majors' Arithmetic Misconceptions With Short-Term Instruction Using Manipulatives. *The Journal of Educational Research*, *101*(4), 234–242. https://doi.org/10.3200/JOER.101.4.234-242

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D., Cushman, F., Stewart, L., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, *293*(5537), 2105. https://doi.org/10.1126/science.1062872

Gunnell, J. J., & Ceci, S. J. (2010). When emotionality trumps reason: A study of individual processing style and juror bias. *Behavioral Sciences & the Law*, *28*(6), 850–877. https://doi.org/10.1002/bsl.939

Hart, W., Kinrade, C., Tortoriello, G. K., Richardson, K., & Breeden, C. J. (2020). Identifying with the "bad" guy: A novel account for apparent moral-judgment deficiencies in

antagonistic personality. *Personality and Individual Differences*, *166*, 110161. https://doi.org/10.1016/j.paid.2020.110161

Hechler, S., & Kessler, T. (2018). On the difference between moral outrage and empathic anger: Anger about wrongful deeds or harmful consequences. *Journal of Experimental Social Psychology*, *76*, 270–282. https://doi.org/10.1016/j.jesp.2018.03.005

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598–603. https://doi.org/10.1016/j.paid.2012.11.008

Kang, J., Bennett, M., Carbado, D., & Casey, P. (2012). Implicit bias in the courtroom. *UCLA Law Review*, *59*(5), 1124–1187.

Kaplan, M. F., & Miller, L. E. (1978). Reducing the effects of juror bias. *Journal of Personality and Social Psychology*, *36*(12), 1443–1455. https://doi.org/10.1037/0022-3514.36.12.1443

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

Kroneisen, M., & Heck, D. W. (2020). Interindividual differences in the sensitivity for consequences, moral norms, and preferences for inaction: Relating basic personality traits to the CNI model. *Personality and Social Psychology Bulletin*, *46*(7), 1013-1026.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lane, R. D., Weihs, K. L., Herring, A., Hishaw, A., & Smith, R. (2015). Affective agnosia: Expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy. *Neuroscience & Biobehavioral Reviews*, *55*, 594–611. https://doi.org/10.1016/j.neubiorev.2015.06.007

Lee, K., & Ashton, M. (2008). The HEXACO personality factors in the indigenous personality lexicons of English and 11 other languages. *Journal of Personality*, *76*(5), 1001–1054.

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences*, *11*(8), 319–321. https://doi.org/10.1016/j.tics.2007.06.001

Moll, J., de Oliveira-Souza, R., & Eslinger, P. J. (2003). Morals and the human brain: A working model. *Neuroreport*, *14*(3), 299–305. https://doi.org/10.1097/00001756-200303030-00001

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences*, *108*(7), 2688. https://doi.org/10.1073/pnas.1011734108

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*(3), 530–542. https://doi.org/10.1016/j.cognition.2005.07.005

Patil, I., & Silani, G. (2014a). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00501

Patil, I., & Silani, G. (2014b). Alexithymia increases moral acceptability of accidental harms. *Journal of Cognitive Psychology*, *26*(5), 597–614. https://doi.org/10.1080/20445911.2014.929137

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2021). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology*, *120*(2), 443–460. https://doi.org/10.1037/pspp0000281

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

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*(2), 242–258. https://doi.org/10.1037/0022-3514.51.2.242

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17*(4), 251–283. https://doi.org/10.1002/per.474

Pfattheicher, S., Sassenrath, C., & Keller, J. (2019). Compassion magnifies third-party punishment. *Journal of Personality and Social Psychology*, *117*(1), 124–141. https://doi.org/10.1037/pspi0000165

Reynolds, S. J. (2008). Moral attentiveness: Who pays attention to the moral aspects of life? *Journal of Applied Psychology*, *93*(5), 1027–1041. https://doi.org/10.1037/0021-9010.93.5.1027

Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is conscientiousness and how can it be assessed? *Developmental Psychology*, *50*(5), 1315–1330. https://doi.org/10.1037/a0031109

Schriber, R. A., Robins, R. W., & Solomon, M. (2014). Personality and self-insight in individuals with autism spectrum disorder. *Journal of Personality and Social Psychology*, *106*(1), 112–130. PubMed. https://doi.org/10.1037/a0034950

Shaffer, D. R., & Case, T. (1982). On the decision to testify in one's own behalf: Effects of withheld evidence, defendant's sexual preferences, and juror dogmatism on juridic decisions. *Journal of Personality and Social Psychology*, *42*(2), 335–346. https://doi.org/10.1037/0022-3514.42.2.335

Shaffer, D. R., Plummer, D., & Hammock, G. (1986). Hath he suffered enough? Effects of jury dogmatism, defendant similarity, and defendant's pretrial suffering on juridic decisions. *Journal of Personality and Social Psychology*, *50*(5), 1059–1067. https://doi.org/10.1037/0022-3514.50.5.1059

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*(1), 76–105. https://doi.org/10.1016/0022-1031(91)90011-T

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454–458. https://doi.org/10.1016/j.cognition.2011.01.018

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, *64*, 231–255.

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

Trémolière, B., Neys, W. D., & Bonnefon, J.-F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*(3), 379–384. https://doi.org/10.1016/j.cognition.2012.05.011

van Doorn, J., Zeelenberg, M., Breugelmans, S. M., Berger, S., & Okimoto, T. G. (2018). Prosocial consequences of third-party anger. *Theory and Decision*, *84*(4), 585–599. https://doi.org/10.1007/s11238-017-9652-6

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, *107*(15), 6753–6758. https://doi.org/10.1073/pnas.0914826107

Young, Liane, Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, *65*(6), 845–851. https://doi.org/10.1016/j.neuron.2010.03.003

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, Act, and Outcome in Behavioral Prediction and Moral Judgment. *Child Development*, *67*(5), 2478–2492. JSTOR. https://doi.org/10.2307/1131635

Zuckerman, M. (2007). *Sensation seeking and risky behavior.* (pp. xix, 309). American Psychological Association. https://doi.org/10.1037/11555-000

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer New York. https://doi.org/10.1007/978-0-387-87458-6