COMPREHENSIVE RESULTS IN SOCIAL PSYCHOLOGY
https://doi.org/10.1080/23743603.2022.2102473

Check for updates

Prediction of dishonest behavior using the CNI model of moral dilemma judgment: a registered report

Nyx L. Ng, Dillon M. Luke and Bertram Gawronski

Department of Psychology, University of Texas, Austin, TX, USA

ABSTRACT

Drawing on moral philosophy, research in moral psychology has used hypothetical sacrificial dilemmas to understand how moral judgments are made by laypeople. Although heavily influential, a frequent question raised is whether responses to hypothetical scenarios are informative about instances of morally relevant behavior. Using the CNI model to quantify sensitivity to consequences, sensitivity to moral norms, and general preference for inaction versus action in responses to sacrificial dilemmas, two preregistered studies examined whether specific factors underlying sacrificial dilemma judgments predict dishonest behavior for personal monetary gain (Study 1) and the greater good (Study 2). Testing the hypothesis that specific factors underlying sacrificial dilemma judgments are linked to other morally relevant behaviors via broader underlying moral dispositions, confirmatory tests provide strong support the predicted associations between sensitivity to moral norms and dishonest behavior. No support was found for a predicted association between sensitivity to consequences and dishonest behavior for the greater good. The findings contribute to ongoing debates about the value of individual differences in sacrificial dilemma judgments for predicting instances of morally relevant behavior.

ARTICLE HISTORY
Received 22 May 2021
Accepted 21 June 2022

KEYWORDS
CNI model; deontology; dishonest behavior; moral judgment; utilitarianism

Hypothetical sacrificial dilemmas have long been used in moral philosophy as a way of evaluating normative theories about how moral judgments *ought to be made* (Foot, 1967; Thomson, 1985). Drawing on this tradition, moral psychology has used these dilemmas to inform descriptive theories of how moral judgments are *actually made* by laypeople (e.g. Cushman, 2013; Greene, 2007). Although research in the latter tradition has provided valuable insights into the mental underpinnings of sacrificial dilemma judgments, this body of work has also been the subject of criticism (Hester & Gray, 2020; Schein, 2020). While some concerns have focused on the failure of sacrificial dilemmas to capture important aspects of moral judgment (e.g. Kahane et al., 2018), others have raised questions about whether responses to sacrificial dilemmas are related to instances of morally relevant behavior (Bauman et al., 2014).

The primary goal of the current research was to provide insight into the latter question by investigating the relation between sacrificial dilemma judgments and dishonest behavior. In two preregistered studies, we examined whether sacrificial dilemma judgments

CONTACT Nyx L. Ng nyx@utexas.edu Department of Psychology, University of Texas at Austin, 108 East Dean Keeton, A8000, Austin, TX 78712-0187, USA

© 2022 European Association of Social Psychology

predict dishonest behavior for personal monetary gain (Study 1) and the greater good (Study 2). To provide more nuanced insights into associations between sacrificial dilemma judgments and dishonest behavior, we used the CNI model of moral decision-making (Gawronski et al., 2017) to quantify (1) sensitivity to consequences, (2) sensitivity to moral norms, and (3) general preference for inaction versus action in responses to sacrificial dilemmas. Our main question was whether dishonest behavior is systematically related to specific factors captured by the CNI model.

Sacrificial dilemma research

Perhaps the most well-known sacrificial dilemma used in psychological research is the trolley problem, in which a runaway trolley is said to be on a lethal collision course with five railroad workers. In a version called the switch dilemma, participants are told that it would be possible to pull a lever, redirecting the trolley onto another track where it would kill one person instead of the five workers (Foot, 1967). In another version called the footbridge dilemma, participants are told that it would be possible to push a large man in front of the trolley to obstruct its path, killing the man but saving the five workers (Thomson, 1985). Favoring action by either pulling the lever or pushing the man has been described as a characteristically utilitarian judgment in the sense that it maximizes overall outcomes for the greater good (i.e. killing one saves the lives of five; see, Conway et al., 2018). Conversely, opposing action on the two dilemmas has been described as a characteristically deontological judgment in the sense that it conforms to moral norms relevant to the situation (i.e. the moral norm that prohibits the killing of other people; see, Conway et al., 2018).

Using this paradigm, moral psychology has made significant strides in understanding how people make judgments about right and wrong. This work has focused on various contextual influences on sacrificial dilemma judgments (e.g. Suter & Hertwig, 2011; Valdesolo & DeSteno, 2006), individual differences in sacrificial dilemma judgments (e.g. Moore et al., 2011; Patil, 2015), and neural underpinnings of sacrificial dilemma judgments (e.g. Greene et al., 2001; Koenigs et al., 2007). Informed by the findings of this work, several theories have been developed that aim at identifying the mental processes underlying sacrificial dilemma judgments (e.g. Crockett, 2013; Cushman, 2013; Greene, 2007).

While the sacrificial dilemma paradigm has inspired a substantial amount of research, it has also drawn criticism. One frequent criticism concerns the artificial nature of the scenarios used in sacrificial dilemma research. Some researchers noted that participants tend to perceive scenarios like the trolley problem as humorous, unrealistic, and implausible (e.g. Bauman et al., 2014; Körner et al., 2019), raising questions about whether responses to these scenarios involve the same mental processes that underlie naturally occurring instances of morally relevant behavior. If sacrificial dilemma judgments are indeed shaped by different processes, they may be of limited value for understanding instances of morally relevant behavior in real-world contexts. Addressing these concerns, a small but growing body of research has examined the utility of sacrificial dilemma judgments in predicting instances of morally relevant behavior.

Sacrificial dilemma judgment and moral behavior

One line of research has examined the relation between sacrificial dilemma judgments and behaviors in parallel real-world situations involving trade-offs of the welfare of one against the many. In an influential study, Bostyn et al. (2018) examined whether preference for utilitarian judgments on hypothetical sacrificial dilemmas was predictive of behavioral choices on a "real-life" sacrificial dilemma involving the decision of whether to deliver an electric shock to one mouse to prevent the delivery of electric shocks to five mice. Despite the conceptual similarity between the two measures, preference for utilitarian judgments on hypothetical dilemmas was found to be unrelated to behavioral choices on the "real-life" dilemma. Yet, further exploratory analyses indicated that both preference for deontological judgments and the relative preference for utilitarian over deontological judgments (calculated as a difference score) on hypothetical dilemmas were marginally predictive of behavioral choices on the "real-life" dilemma (Plunkett & Greene, 2019). Based on these latter findings, some have interpreted the results of this study to be inconclusive (Plunkett & Greene, 2019; but see, Bostyn & Roets, 2019).

In a related study, Bostyn et al. (2019) examined whether judgments on hypothetical sacrificial dilemmas were predictive of behavioral choices on a monetary sacrificial dilemma involving the decision of whether to take away a portion of the compensation awarded to one relatively well-compensated participant to be redistributed to five other relatively poorly compensated participants.[1] In a meta-analysis across four studies, preference for utilitarian judgments on hypothetical dilemmas was found to positively predict reallocating money, providing support for a relation between sacrificial dilemma judgment and instances of morally relevant behavior.

In another line of work, studies have examined whether sacrificial dilemma judgments are predictive of choices in economic games. Across this research, sacrificial dilemma judgments were not significantly associated with choices on a dictator game (Capraro et al., 2018) and a trust game (Bostyn & Roets, 2017; Capraro et al., 2018). However, deontological inclinations in sacrificial dilemmas have shown a positive association with cooperative behavior in a public goods game (Bostyn & Roets, 2017), and morally dubious judgments on sacrificial dilemmas (e.g. actively causing death without saving additional lives) have shown a positive association with "money burning" or the willingness to reduce the amount of money given to another person at cost to oneself (Dickinson & Masclet, 2019). Taken together, findings across studies have been mixed and inconclusive, leaving the question of whether sacrificial dilemma judgments predict instances of morally relevant behavior up for debate.[2]

The current research aims to shed light on this question by building on past research in three notable ways. First, the current research investigated a novel instance of morally relevant behavior in the form of dishonest behavior. Some prior studies examined behavioral choices on sacrificial dilemmas mirroring the structure of the trolley problem (e.g. Bostyn & Roets, 2019; Bostyn et al., 2018). An advantage of this approach is that it guarantees a high degree of correspondence between moral judgment and behavior, which should maximize potential associations between the two (Ajzen & Fishbein, 1977). However, a downside of this approach is that the criticism regarding the artificiality of judgments on hypothetical sacrificial dilemmas may also be levied against behavioral choices in conceptually similar dilemmas. Moreover, the value of individual

differences in sacrificial dilemma judgments would be much greater if these differences were predictive of a wider range of morally relevant behaviors beyond those described in the scenarios. Such predictive associations would go beyond the question of whether people simply do what they say, suggesting a role of broader underlying moral dispositions that guide not only sacrificial dilemma judgments but a wider range of morally relevant behaviors. Based on these considerations, the current work follows past research in examining behaviors beyond the context of sacrificial dilemmas (Bostyn & Roets, 2017; Capraro et al., 2018; Dickinson & Masclet, 2019), while extending this research by examining dishonest behavior as a yet untested instance of morally relevant behavior. Although this approach carries with it a lower correspondence between moral judgment and behavior, the behavioral criterion should afford better generalizability to naturalistic behavior, given that dishonest behavior is frequently occurring in everyday life (Hofmann et al., 2014).

Second, the current research builds on past research by improving the reliability of measurement. As noted by Dang et al. (2020), one reason behavioral measures can show weak associations with self-report measures is that they suffer from poor reliability. Because most of the obtained null effects in this line of work come from studies relying on a single response to sacrificial dilemmas or a single response in economic games (Bostyn & Roets, 2017; Bostyn et al., 2018; Capraro et al., 2018) and because measures involving a single observation are especially prone to measurement error (Dang et al., 2020), it is possible that the results of these studies underestimate associations between sacrificial dilemma judgments and behavior. To address this concern, the current research assessed moral judgments and dishonest behavior using measures that aggregate responses over many trials. In doing so, measurement error should be minimized, which should increase the power to detect potential associations.

Third, the current work extends past research by using a formal modeling approach to disentangle multiple distinct factors in the measurement of sacrificial dilemma judgments. A considerable amount of past research has used scenarios similar in structure to the trolley dilemma (for an exception, see, Bostyn & Roets, 2017). A problem with this practice is that it includes two major confounds. First, utilitarian and deontological responses are pit against one another, thereby confounding maximization of outcomes and adherence to moral norms in the measurement of moral judgments (Conway & Gawronski, 2013). Second, utilitarian judgments are typically conflated with action while deontological judgments are conflated with inaction, confounding the two kinds of judgments with general action tendencies (Crone & Laham, 2017). Together, these concerns point to three factors that can produce differences in sacrificial dilemma judgments: (1) differences in sensitivity to consequences, (2) differences in sensitivity to moral norms, and (3) differences in general preference for inaction versus action. In the current research, we used a formal modeling approach to disentangle the three factors, which allowed us to investigate their unique relations with dishonest behavior. Moreover, because confounding multiple factors can suppress existing associations between a given factor and a relevant criterion, separating the three factors may allow us to detect associations that are difficult to detect with the traditional dilemma approach.

The CNI model

The CNI model of moral decision-making (Gawronski et al., 2017) is a multinomial model that separately quantifies sensitivity to consequences, sensitivity to moral norms, and general preference for inaction versus action in responses to sacrificial dilemmas. To this end, the CNI model utilizes responses to four types of dilemmas that vary in terms of consequences for the greater good (i.e. the focal action produces benefits that are either greater or smaller than the costs) and relevant moral norms (i.e. the focal action is either prohibited by a proscriptive norm or prescribed by a prescriptive norm). An example dilemma scenario in its four variants is presented in Table 1. As depicted in the processing tree in Figure 1, each factor is captured by a parameter that is characterized by a unique pattern of responding across the four types of dilemmas. Sensitivity to consequences is captured by the *C* parameter, which reflects the extent to which actions are favored when their benefits outweigh their costs but opposed when their costs outweigh their benefits (first row in Figure 1). Sensitivity to moral norms is captured by the *N* parameter, which reflects the extent to which actions are favored when they are prescribed by a prescriptive norm and opposed when they are prohibited by a proscriptive norm (second row in Figure 1). General preference for inaction versus action is captured by the *I* parameter, which reflects the extent to which actions are generally opposed (third row in Figure 1) or generally favored (fourth row in Figure 1).

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 et al. (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?

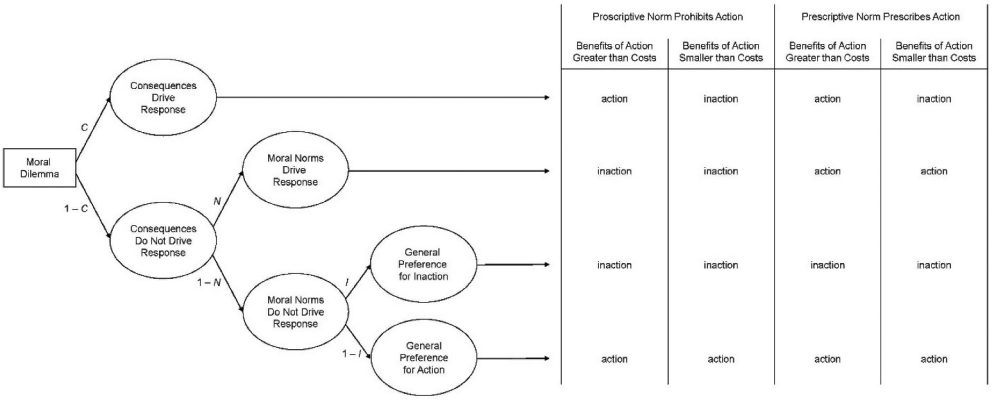

Figure 1. CNI model of moral decision-making predicting action versus inaction responses in moral dilemmas with proscriptive and prescriptive norms and consequences involving benefits of action that are either greater or smaller than costs of action. Reproduced from Gawronski et al. (2017). Reprinted with permission from the American Psychological Association.

Disentangling the three factors underlying sacrificial dilemma judgments has the potential to provide more nuanced insights into associations between sacrificial dilemma judgments and dishonest behavior compared with the traditional dilemma approach. In support of this possibility, research using the CNI model has provided valuable insights into complex associations between sacrificial dilemma judgments and psychopathy. When using the traditional approach to measuring sacrificial dilemma judgment, associations between sacrificial dilemma judgments and psychopathy have been mixed, with a recent meta-analysis suggesting a small positive association between psychopathy and preference for utilitarian over deontological judgments across studies (Marshall et al., 2018). Yet, research using the CNI model revealed a more complex pattern, showing that psychopathy is negatively associated with all three model parameters (Gawronski et al., 2017; Körner et al., 2020; Luke & Gawronski, 2021b). Specifically, participants high (vs. low) in psychopathy showed (1) a weaker sensitivity to consequences, (2) a weaker sensitivity to moral norms, and (3) a weaker general preference for inaction versus action. While the observed associations were small-to-medium in size for sensitivity to consequences and general preference for inaction versus action, the observed associations with sensitivity to moral norms qualify as large in terms of current conventions (Cohen, 1988). These findings suggest that individuals with elevated psychopathic traits (1) are less (not more) utilitarian than others and (2) show much larger differences in moral judgment than would be anticipated based on past work using the traditional dilemma approach. By separately quantifying the three factors underlying sacrificial dilemma judgments, analyses using the CNI model may similarly provide more nuanced insights into the association between sacrificial dilemma judgments and dishonest behavior.

The current research

The primary goal of the current research was to provide insight into the question of whether sacrificial dilemma judgments are predictive of instances of morally relevant behavior. In two preregistered studies, we examined the extent to which differences in

the resolution of sacrificial dilemmas are associated with differences in dishonest behavior, a morally relevant behavior with high real-world relevance. To disentangle the different factors contributing to sacrificial dilemma judgments, we used the CNI model to analyze sacrificial dilemma judgments (Gawronski et al., 2017). To measure dishonest behavior with a high degree of reliability, we utilized a coin-toss prediction task that provides participants the repeated opportunity to improve their task performance by lying (adapted from Greene & Paxton, 2009; Shalvi & De Dreu, 2014). In Study 1, better performance in the coin-toss prediction task led to increased personal monetary gain. In Study 2, better performance in the coin-toss prediction task led to increased benefits for the greater good.

To permit comparisons with past research, we first performed exploratory analyses regarding the correlations between dishonest behavior and preference for utilitarian over deontological judgments, with the latter being conceptualized as the preference for action (vs. inaction) on dilemmas where action is prohibited by a proscriptive norm and action leads to better outcomes for the greater good (similar to the trolley problem). Expanding on our exploratory analyses, we conducted confirmatory analyses testing specific predictions about correlations between dishonest behavior and the three CNI parameters. To test the robustness of the results of the correlation analyses, we conducted confirmatory follow-up analyses, testing the predicted associations using multiple-regression analyses in which dishonest behavior was simultaneously regressed onto all three parameters of the CNI model.

To confirm the quality of our measures of moral judgment and dishonest behavior for interpretations of potential null effects, we analyzed the internal consistencies of the predictor and criterion measures and included two positive-control measures at the end of each study. First, participants were asked to complete Levenson, Kiehl, and Fitzpatrick's (1995) Primary Psychopathy Scale (PPS). Six independent studies from our group obtained reliable negative associations between PPS scores and the three parameters of the CNI model, with average correlations of $r = -.27$ for sensitivity to consequences, $r = -.53$ for sensitivity to moral norms, and $r = -.20$ for general preference for inaction versus action (Gawronski et al., 2017; Körner et al., 2020; Luke & Gawronski, 2021b).[3] We expected to replicate these associations in the current studies, which would confirm the quality of our moral judgment measures for the interpretation of potential null effects regarding the association between moral judgment and dishonest behavior. Second, participants were asked to complete the honesty–humility subscale of the HEXACO-60 inventory (Ashton & Lee, 2009). Previous research revealed reliable associations between individual differences in honesty–humility and dishonest behavior on coin-toss prediction tasks similar to the one in the current studies (e.g. Heck et al., 2018; Hilbig & Zettler, 2015). These associations have been found to be medium-to-large in size and to replicate across multiple task contexts. We expected to replicate these associations in the current studies, which would confirm the quality of our measure of dishonest behavior for the interpretation of potential null effects regarding the association between moral judgment and dishonest behavior.

For each study, we aimed to obtain a sample of approximately 258 participants, which provides a statistical power of 90% in detecting a small-to-medium correlation of $r = .20$ ($\alpha = .05$; two-tailed).[4] In addition to exceeding the sample size required for stable estimates of the magnitude of a correlation ($N = 250$; see, Schönbrodt & Perugini, 2013),

this sample size also provided a statistical power of 90% in detecting a small-to-medium effect of $f^2 = .04$ in multiple regression analyses assuming one tested predictor and three total predictors ($\alpha = .05$; two-tailed). Taken together, the sample sizes planned for the two studies were well powered to detect theoretically meaningful associations between moral judgment and moral behavior.

Based on past research in our laboratory using similar procedures, we expected that data from 10% to 15% of the sample would be excluded from analyses based on three preregistered exclusion criteria. Specifically, participants were excluded from analyses if they (1) failed an instructional attention check, (2) responded with the same key on all sacrificial dilemmas, or (3) showed accuracy rates that fell below 40% in the coin-toss prediction task. Thus, to achieve our sample-size goal, we oversampled and recruited 300 participants for each study and conducted all analyses with the remaining sample after applying our preregistered exclusion criteria. We report all measures, conditions, and data exclusions. The data, analysis codes, and materials for the two studies can be accessed at https://osf.io/3c6dh/. The two studies were preregistered prior to data collection at https://osf.io/zktrf.

Study 1

Study 1 investigated associations between sacrificial dilemma judgments and dishonest behavior when dishonest behavior led to increased personal monetary gain. Toward this end, participants first completed a battery of sacrificial dilemmas for research using the CNI model (Körner et al., 2020). Afterward, participants completed a coin-toss prediction task designed to measure dishonest behavior (Greene & Paxton, 2009; Shalvi & De Dreu, 2014). On each trial of the coin-toss prediction task, participants were asked to (1) make a prediction about a coin toss, (2) observe a simulated coin toss on the computer screen, and (3) indicate whether their prediction was correct or incorrect. Participants were told that the three participants with the best performance on the task would each receive a bonus payment of $50 in addition to their compensation for participating in the study. Because participants did not have to record their predictions, they could increase their chance of winning a bonus payment by lying about their performance. With a total of 70 prediction trials and two potential outcomes (i.e. heads vs. tails), accuracy scores (i.e. number of correct predictions) exceeding 35 correct predictions are increasingly improbable and suggestive of dishonest reporting. Our main question was whether sacrificial dilemma judgments are systematically related to dishonest behavior on the coin-toss prediction task.

To permit comparisons with past research, we first performed an exploratory analysis regarding the correlation between dishonest behavior and preference for utilitarian over deontological judgments, with the latter being conceptualized as preference for action (vs. inaction) on dilemmas where action is prohibited by a proscriptive norm and action leads to better outcomes for the greater good (similar to the trolley problem).[5] Expanding on this exploratory analysis, we conducted confirmatory analyses testing zero-order correlations between dishonest behavior and each of the three CNI parameters, followed by multiple regression analyses in which dishonest behavior was simultaneously regressed onto all three parameters of the CNI model.

The *N* parameter reflects a behavioral tendency to conform to relevant moral norms in responses to sacrificial dilemmas. Based on the propositions that (1) this tendency reflects a broader disposition to conform to relevant moral norms in other situations and (2) lying to enhance task performance for personal monetary gain would constitute a violation of moral norms surrounding honesty, we expected that participants who show greater sensitivity to moral norms in their sacrificial dilemma judgments would be less likely to lie in the coin-toss prediction task. This association should be reflected in a significant negative correlation between the CNI model's *N* parameter and accuracy scores in the coin-toss prediction task (Hypothesis 1a). We further predicted that the *N* parameter would show a significant negative association with accuracy scores in the multiple regression analysis controlling for the *C* and *I* parameters (Hypothesis 1b). Because dishonest behavior in Study 1 served personal monetary gains rather than the greater good and there is no theoretical reason to expect a link between dishonest behavior and general action tendencies, we did not have any specific hypotheses regarding associations with either the *C* or *I* parameters. For the sake of comprehensiveness, we examined these associations in an exploratory manner.

Methods

Participants

A total of 300 participants were recruited using Prolific, a crowdsourcing platform for online research (Peer et al., 2017). The data collection stopped once 300 participants were approved for compensation. Eligibility for participation was restricted to Prolific workers who (1) were United States nationals, (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' laboratory using the sacrificial dilemmas used in the current research. Completion of the study took approximately 45 minutes, and participants were compensated $6.00 for their time.

Out of the 300 participants who completed the study, 22 failed the attention check and another five had accuracy scores of less than 40% for the coin-toss prediction task. The final sample for Study 1 was thus $N = 273$. Participant demographics are as follows: 62.3% female, 33.0% male, 3.3% other, and 1.5% preferred not to answer; 75.8% identified as White, 12.5% Black/African American, 8.8% Asian, 3.3% American Indian/Alaska Native, and 6.6% other races; $M_{age} = 36.8$ years, $SD_{age} = 12.7$.

Procedure and materials

After providing informed consent, participants were asked to complete a validated battery of 44 sacrificial dilemmas for research using the CNI model (Körner et al., 2020). The battery comprised 11 basic scenarios, each with four variants manipulating consequences (i.e. the described action produces benefits that are either greater versus smaller than the costs) and moral norms (i.e. the described action is either prescribed or prohibited by a moral norm).[6] For each dilemma, participants were asked to indicate whether they would perform the described action using the response options *yes* vs. *no*. Dilemmas were presented in a predetermined random order that was fixed for all participants.

After completing the dilemma battery, participants were informed that they are beginning the second part of the study. Participants were told that the task in the second part was designed to explore the claims of a controversial study by Bem (2011), suggesting that people may be able to predict the outcomes of random events before they happen. Participants then completed a coin-toss prediction task designed to measure dishonest behavior (Greene & Paxton, 2009; Shalvi & De Dreu, 2014). On each trial of the task, participants were instructed to make a prediction about the outcome ("heads" vs. "tails"), watch a simulated coin toss on the computer screen, and then report whether their initial prediction was correct or incorrect. Participants were informed that the three people with the highest number of correct predictions at the end of the study will each receive a $50 bonus payment.[7] Participants were asked to make predictions about the outcomes of 70 coin tosses, each of which was programmed to randomly result in either "heads" or "tails." Because initial predictions were not recorded, participants were able to lie about their performance by reporting incorrect predictions as correct. To avoid fatigue, the task was divided into two blocks of 35 trials, with a short break after the first 35 trials. At the end of the task, participants were provided with feedback regarding their final performance.

After completing the coin-toss prediction task, participants completed two positive-control measures: Levenson et al.'s (1995) PPS and the honesty–humility subscale of the HEXACO-60 inventory (Ashton & Lee, 2009). Following Levenson et al. (1995), responses to the 16 items of the PPS were measured with 4-point rating scales with the response options *disagree strongly* (1), *disagree somewhat* (2), *agree somewhat* (3), and *agree strongly* (4). Following Ashton and Lee (2009), responses to the 10 items of the honesty–humility subscale were measured with 5-point rating scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

After completing the two positive-control measures, participants answered a set of demographics questions, completed a reading-intensive attention check (see below), were debriefed, and then redirected for payment of their compensation. The $50 bonus payments to the three participants with the highest accuracy scores on the coin-toss prediction task were made through Prolific after completion of the data collection. To conform with IRB requirements, identifying information for the payment of the bonus (i.e. Prolific ID) was deleted from the data files after the bonus had been granted.

Data aggregation

To permit a comparison of our findings with past research, we calculated a score reflecting the relative preference for action (vs. inaction) on sacrificial dilemmas where action is prohibited by a proscriptive norm and action leads to better outcomes for the greater good (similar to the trolley problem). Toward this end, we calculated the proportion of responses endorsing action (*yes*) versus inaction (*no*) on dilemmas of this type, which resulted in an aggregate index we refer to as the traditional score. Higher scores on this index can be interpreted as reflecting a greater relative preference for utilitarian over deontological judgments (see, Conway et al., 2018).

Scores on the three CNI model parameters were estimated for each participant following the procedures described by 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. Based on the processing tree depicted in Figure 1, it is possible to derive 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 Appendix).[8] Using maximum likelihood statistics, it is possible to estimate specific values for the three model parameters, such that the discrepancy between the estimated probability of action (vs. inaction) responses across sacrificial dilemmas and the observed probability of action (vs. inaction) responses across sacrificial 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. The resulting model had three free parameters (i.e. *C*, *N*, and *I*) and four free response categories (i.e. four kinds of dilemmas), resulting in one degree of freedom. Following Gawronski et al. (2017), the 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) at https://osf.io/ndf4w/. Research by Luke and Gawronski (2021c) obtained high reliability estimates for the *C* and the *N* parameter in terms of their internal consistency (α's > .69) and test–retest correlations (*r*s > .80); reliability estimates obtained for the *I* parameter were found to be lower for both internal consistency (Cronbach's αs between .37 and .53) and test–retest stability ($r = .41$). To confirm the reliability of the three 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, Gawronski et al., 2020; Luke & Gawronski, 2021c). The internal consistencies of the three parameters were estimated by calculating a Cronbach's α value for each parameter based on the two scores.

Participants' accuracy scores on the coin-toss prediction task were used as an index of dishonest behavior. With a total of 70 prediction trials and two potential outcomes (i.e. heads vs. tails), participants should be able to predict the correct outcome of approximately 35 of the 70 trials by mere chance. Thus, accuracy scores exceeding 35 correct predictions are increasingly improbable and suggestive of dishonest responding. Likewise, accuracy scores lower than 35 correct guesses are increasingly improbable and suggestive of dishonest responding. While we did not expect participants to underreport their accuracy, underreporting one's accuracy would not represent dishonest behavior for personal monetary gain. Thus, to obtain a clean measure of dishonest behavior committed specifically for personal monetary gain, we excluded data from participants with accuracy scores less than 28 (40%; Binomial $P = 0.036$) from the analyses. To confirm the reliability of the criterion measure, we estimated two scores of dishonest behavior for each participant, one based on coin-toss trials with odd item numbers and one based on coin-toss trials with even item numbers. The internal consistency of the measure was estimated by calculating a Cronbach's α value based on the two scores. Responses on the two positive-control measures were aggregated by reverse coding negatively framed items and calculating average scores across all items of each scale. The internal consistencies of the measures were estimated by calculating Cronbach's α values across all items.

Missing data and data exclusions

Participants who terminated the study prior to completing all components did not receive compensation. Data from these participants were excluded from analyses. Because participants could not skip responses, we did not anticipate missing values in the remaining data set. Participants with complete data were excluded from analyses if they (1) failed an instructional attention check, (2) responded with the same key on all 44 sacrificial dilemmas, or (3) showed accuracy scores less than 40% (28 trials) on the coin-toss prediction task. 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 question without selecting any answer choices. The attention check includes the following instructions:

> *To facilitate our research on decision-making, we are interested in learning a little more about you, the decision-maker. Psychological research using text-based materials requires that study participants read the materials and do not skip over longer pieces of text. We are therefore interested in whether you actually take the time to read the directions; if not, then some of our manipulations that rely on changes in the instructions will be ineffective. To demonstrate that you have read the instructions, please ignore the question below and all of the response options. Instead, simply continue on to the next page without answering the question. Thank you very much.*

The instructions were followed by the question: Of the following destinations, which one would be your first choice for a vacation if you had a free all-inclusive round trip after the Covid-19 pandemic? With the response options: *Australia, Brazil, China, Egypt, France, Germany, India, Japan, New Zealand, Mexico, Russia, South Africa, Spain, Sweden,* and *the United Kingdom*. If a participant answered the question by selecting any of the 15 response options, the participant failed to correctly follow instructions and was therefore excluded from analyses.

Analysis plan

For the sake of robustness, we employed a nonparametric bootstrapping method for all exploratory and confirmatory correlational and regression analyses. This approach involved (1) obtaining k bootstrap samples by resampling from the full sample with replacement, (2) calculating the relevant sample statistic (i.e. correlation coefficient and beta coefficients) in each of k samples, (3) using the resulting sample statistics obtained across k samples to form a distribution, and (4) constructing a confidence interval based on the 2.5th and 97.5th percentiles in this distribution. Effects were interpreted as statistically significant if the bootstrapped confidence interval for a given effect excludes 0. These analyses were conducted using the boot package in R (Canty & Ripley, 2021), and we obtained 10,000 bootstrap samples for each analysis.

To permit comparisons of our findings with past research, initial exploratory analyses examined the correlation between accuracy scores in the coin-toss prediction task and the traditional score of preference for utilitarian over deontological judgments. Confirmatory analyses investigated associations between dishonest behavior and the three parameters of the CNI model (Gawronski et al., 2017). First, we separately tested

the zero-order correlations between accuracy scores in the coin-toss prediction task and the *C* parameter (sensitivity to consequences), the *N* parameter (sensitivity to moral norms), and the *I* parameter (general preference for inaction versus action). Second, we conducted multiple-regression analyses, simultaneously regressing accuracy scores onto the three parameters of the CNI model. To assess the quality of our measures of moral judgment and moral behavior for interpretations of potential null effects, we further examined (1) zero-order correlations between PPS scores and the three CNI parameters and (2) zero-order correlations between scores on the honesty–humility subscale and accuracy scores (see above).

Results: preregistered analyses

Table 2 presents the means, standard errors, correlation coefficients, and bootstrapped 95% confidence intervals for the three CNI model parameters, accuracy scores on the coin-toss prediction task, and the positive control measures. The mean and median scores for the coin-toss prediction task were 42 and 39 accurate guesses, respectively. Eight participants reported a 100% accuracy rate. In terms of internal consistency, Cronbach's α for the *C, N*, and *I* parameters were .58, .68, and .43, respectively. Cronbach's α for the coin-toss accuracy scores, PPS, and honesty–humility scale were higher at .82, .89, and .83, respectively.

Traditional score

Initial exploratory analyses revealed that accuracy on the coin-toss prediction task was not significantly associated with preference for utilitarian over deontological judgments captured by the traditional score, $r = .11$, 95% CI [−.02, .24].

CNI parameters

Supporting Hypothesis 1a, confirmatory analyses revealed a significant negative correlation between participants' accuracy on the coin-toss prediction task and the *N* parameter, $r = -.14$, 95% CI [−.26, −.02], indicating that participants who were more sensitive to moral norms were less likely to cheat on the coin-toss prediction task. Further exploratory analyses revealed no significant associations between accuracy on the coin-toss prediction task and both the *C* parameter, $r = -.04$, 95% CI [−.16, .09], and *I* parameter, $r = -.03$, 95% CI [−.15, .08].

To test whether the association between the *N* parameter and accuracy on the coin-toss prediction task remains robust when controlling for the *C* and *I* parameters, we simultaneously regressed coin-toss accuracy scores onto the *C, N*, and *I* parameters. Table 3 presents the results of the multiple-regression analysis. Supporting Hypothesis 1b, the relationship between the *N* parameter and accuracy scores on the coin-toss prediction task remained statistically significant after controlling for the *C* and *I* parameters, $B = -4.94$, 95% CI [−8.99, −1.02].

Table 2. Descriptive statistics, correlation coefficients, and bootstrapped 95% confidence intervals of the traditional dilemma score, CNI model parameters, coin-toss accuracy score, and positive control measures (Study 1).

	Correlation Coefficient [Bootstrapped 95% CI]						
	Traditional score	*C* parameter	*N* parameter	*I* parameter	Coin-toss accuracy	Psychopathy	Honesty–humility
Traditional score	-						
C parameter	.66 [.57, .73]	-					
N parameter	−.66 [−.73, −.57]	−.08 [−.20, .04]	-				
I parameter	−.29 [−.40, −.17]	.15 [.02, .28]	.28 [.17, .41]	-			
Coin-toss accuracy	.11 [−.02, .24]	−.04 [−.16, .09]	−.14 [−.26, −.02]	−.03 [−.15, .08]	-		
Psychopathy	.27 [.15, .38]	−.07 [−.20, .05]	−.46 [−.56, −.35]	−.21 [−.33, −.10]	.12 [−.03, .25]	-	
Honesty–humility	−.31 [−.42, −.20]	−.07 [−.20, .05]	.46 [.36, .56]	.10 [−.03, .22]	−.17 [−.30, −.04]	−.71 [−.77, −.64]	-
Descriptive statistics							
Mean	3.76	.25	.62	.70	41.62	1.63	3.52
SE	.14	.01	.02	.02	.59	.03	.05
95% CI	[3.48, 4.03]	[.23, .27]	[.59, .66]	[.67, .74]	[40.46, 42.77]	[1.57, 1.68]	[3.43, 3.62]
Cronbach's α	-	.58	.68	.43	.82	.89	.83

Table 3. Results of multiple-regression analyses regressing coin-toss accuracy scores onto CNI model parameters.

	Study 1			Study 2		
	B	SE	Bootstrapped 95% CI	*B*	SE	Bootstrapped 95% CI
CNI model parameters						
C parameter	−3.17	3.57	[−10.65, 4.15]	−3.72	2.38	[−8.07, 0.53]
N parameter	−4.94	2.07	[−8.99, −1.02]	−3.14	1.41	[−5.69, −0.67]
I parameter	0.54	2.09	[−3.49, 4.48]	−0.53	1.42	[−2.89, 1.61]
Adj. R^2	.01			.02		

Positive control measures

As expected, scores on the HEXACO-60 honesty–humility scale showed a significant negative correlation with coin-toss accuracy scores, $r = -.17$, 95% CI [−.30, −.04], such that participants who scored higher on the measure of honesty–humility were less likely to cheat on the coin-toss prediction task. Although the relations between primary psychopathy measured with the PPS and the three CNI parameters were all in the expected negative direction, the correlations were statistically significant only for the *N* parameter, $r = -.46$, 95% CI [−.56, −.35], and the *I* parameter, $r = -.21$, 95% CI [−.33, −.10], but not the *C* parameter, $r = -.07$, 95% CI [−.20, .05].

Discussion

Confirmatory correlational and multiple-regression analyses found support for both Hypotheses 1a and 1b. In a situation where dishonest behavior led to potential personal gains, participants who were more sensitive to moral norms were less likely to behave dishonestly by exaggerating their accuracy in predicting the outcomes of the coin toss, and this association remained robust after controlling for sensitivity to consequences and general action tendencies. Exploratory correlational analyses did not reveal any significant associations between dishonest behavior on the coin-toss prediction task and sensitivity to consequences and general action tendencies, respectively. Preference for utilitarian over deontological judgments captured by the traditional dilemma score was also unrelated to dishonest behavior on the coin-toss prediction task.

Study 2

Study 2 investigated associations between sacrificial dilemma judgments and dishonest behavior when dishonest behavior leads to increased benefits for the greater good. Toward this end, participants completed the same battery of 44 sacrificial dilemmas for research using the CNI model (Körner et al., 2020) and the same coin-toss prediction task designed to measure dishonest behavior (Greene & Paxton, 2009; Shalvi & De Dreu, 2014). Yet, different from Study 1, participants in Study 2 were told that we would donate 10¢ to a charity of their choice for every correct prediction on the coin-toss prediction task. Thus, while dishonest behavior led to personal monetary gain in Study 1, dishonest behavior produced benefits for the greater good in Study 2.

Given that (1) scores on the *N* parameter may reflect a broader disposition to conform to relevant moral norms in situations beyond responses to sacrificial dilemmas and (2) lying to enhance task performance would constitute a violation of moral norms surrounding honesty, we expected that participants who show greater sensitivity to moral norms in their sacrificial dilemma judgments would be less likely to lie in the coin-toss prediction task. In line with the predictions in Study 1, this association should be reflected in a significant negative correlation between the CNI model's *N* parameter and accuracy scores in the coin-toss prediction task (Hypothesis 2a) and a significant negative association with accuracy scores in the multiple regression analysis controlling for the *C* and *I* parameters (Hypothesis 2b).

The *C* parameter reflects a behavioral tendency to maximize overall outcomes for the greater good in responses to sacrificial dilemmas. Based on the propositions that (1) this tendency reflects a broader disposition to maximize overall outcomes for the greater good in other situations and (2) lying to enhance task performance leads to increased benefits for the greater good, we expected that participants who showed greater sensitivity to consequences in their sacrificial dilemma judgments would be more likely to lie in the coin-toss prediction task. This association should be reflected in a significant positive correlation between the CNI model's *C* parameter and accuracy scores (Hypothesis 3a) and a significant positive association with accuracy scores in the multiple-regression analysis controlling for the *N* and *I* parameters (Hypothesis 3b). Because there is no theoretical reason to expect a link between dishonest behavior and general action tendencies, we did not have any specific hypotheses regarding associations with the *I* parameter. For the sake of comprehensiveness, we examined this association in an exploratory manner.

Methods

Participants

A total of 300 participants were recruited using Prolific (Peer et al., 2017). The data collection stopped once 300 participants were approved for compensation. The same eligibility criteria for participation used in Study 1 were applied in Study 2. Completion of the study took approximately 45 minutes, and participants were compensated $6.00 for their time.

Out of the 300 participants who completed the study, 24 failed the attention check and another six had accuracy scores that were lower than 40% for the coin-toss prediction task. The final sample for Study 2 was thus $N = 270$. Participant demographics are as follows: 71.5% female, 25.9% male, 2.2% other, and 0.4% preferred not to answer; 81.5% identified as White, 10.4% Black/African American, 8.9% Asian, 1.9% American Indian/Alaska Native, and 4.8% other races; $M_{age} = 34.9$ years, $SD_{age} = 12.8$.

Procedure and materials

The procedure for Study 2 was identical to Study 1 with one notable exception. While dishonest behavior in Study 1 led to personal monetary gain, dishonest behavior in Study 2 produced benefits for the greater good. Toward this end, participants were told that we will donate 10¢ to a charity of their choice for every correct prediction on the coin-toss prediction task. Before starting the coin-toss prediction task, participants were presented

with a list of ten charities (incl. brief descriptions about what their donations would be used to achieve) and asked to choose one charity to which they would like to donate. The list was developed to include charities that span a diverse range of politically neutral causes in the United States, including education, health, and hunger relief. Only charities with a focus on human welfare were considered for inclusion. To emphasize the utilitarian benefits of the donations, we prioritized charities for which relatively small donations could substantively contribute to a concrete impact (e.g. buying school supplies for a student from a poor family). After completion of the study, we sent participants a summary of the combined donations to each charity based on all participants' performance and donated the respective amounts.

Data treatment and analyses

Data aggregation procedures, missing data and data exclusion procedures, and the analytic plan for Study 2 followed the procedures of Study 1.

Results: preregistered analyses

Table 4 presents the means, standard errors, correlation coefficients, and bootstrapped 95% confidence intervals for three CNI model parameters, accuracy scores on the coin-toss prediction task, and the positive control measures. The mean and median scores for the coin-toss prediction task were 39 and 38 accurate guesses, respectively. Three participants reported a 100% accuracy rate. In terms of internal consistency, Cronbach's α were .62, .56, and .39 for the *C, N*, and *I* parameters, respectively. Internal consistency for the coin-toss prediction task as measured using predictions made for even-numbered trials and odd-numbered trials was lower for this study as compared to Study 1 at Cronbach's α = .55. Internal consistencies of the two positive control measures were high at .88 for the PPS and.79 for the HEXACO-60 honesty-humility scale.

Traditional score

Initial exploratory analyses revealed that accuracy on the coin-toss prediction task was not significantly associated with preference for utilitarian over deontological judgments captured by the traditional score, $r = .07$, 95% CI [−.04, .17].

CNI parameters

Supporting Hypothesis 2a, confirmatory analyses revealed a significant negative correlation between accuracy on the coin-toss prediction task and the *N* parameter, $r = -.15$, 95% CI [−.25, −.04], indicating that participants who were more sensitive to moral norms were less likely to cheat by inflating their accuracy on the prediction task. Disconfirming Hypothesis 3a, the correlation between accuracy on the coin-toss prediction task and the *C* parameter was not significant and in a direction that was opposite to the predicted positive association, $r = -.11$, 95% CI [−.22, .01]. There was also no significant correlation between accuracy on the coin-toss prediction task and the *I* parameter, $r = -.07$, 95% CI [−.18, .03].

Table 4. Descriptive statistics, correlation coefficients, and bootstrapped 95% confidence intervals of the traditional dilemma score, CNI model parameters, coin-toss accuracy score, and positive control measures (Study 2).

	Correlation Coefficient [Bootstrapped 95% CI]						
	Traditional score	*C* parameter	*N* parameter	*I* parameter	Coin-toss accuracy	Psychopathy	Honesty–humility
Traditional score	-						
C parameter	.62 [.53, .70]	-					
N parameter	−.58 [−.66, −.49]	.06 [−.07, .19]	-				
I parameter	−.27 [−.37, −.17]	.13 [.01, .25]	.26 [.14, .39]	-			
Coin-toss accuracy	.07 [−.04, .17]	−.11 [−.22, .01]	−.15 [−.25, −.04]	−.07 [−.18, .03]	-		
Psychopathy	.14 [.02, .26]	−.20 [−.32, −.07]	−.36 [−.46, −.25]	−.24 [−.35, −.13]	−.01 [−.13, .10]	-	
Honesty–humility	−.32 [−.42, −.20]	−.08 [−.20, .04]	.38 [.26, .48]	.12 [.01, .24]	−.04 [−.16, .08]	−.62 [−.69, −.54]	-
Descriptive statistics							
Mean	3.94	.25	.59	.69	38.75	1.65	3.50
SE	.13	.01	.02	.02	.40	.03	.04
95% CI	[3.68, 4.20]	[.23, .27]	[.56, .63]	[.66, .73]	[37.95, 39.54]	[1.60, 1.71]	[3.41, 3.59]
Cronbach's α	-	.62	.56	.39	.55	.88	.79

Expanding on the correlation analyses, we simultaneously regressed the coin-toss accuracy scores on the *C, N*, and *I* parameters. Table 3 presents the results of the multiple-regression analysis. Supporting Hypothesis 2b, the relationship between the *N* parameter and accuracy scores on the coin-toss prediction task remained statistically significant after controlling for the *C* and *I* parameters, $B = -3.14$, 95% CI [−5.69, −0.67]. Disconfirming Hypothesis 3b, the *C* parameter was not significantly associated with accuracy scores on the coin-toss prediction task after controlling for the *N* and *I* parameters, $B = -3.72$, 95% CI [−8.07, 0.53].

Positive control measures

Scores on the HEXACO-60 honesty–humility scale were not significantly associated with coin-toss accuracy scores, $r = -.04$, 95% CI [−.16, .08]. Consistent with past research, primary psychopathy measured with the PPS showed significant negative correlations with the *C* parameter, $r = -.20$, 95% CI [−.32, −.07], *N* parameter, $r = -.36$, 95% CI [−.46, −.25], and *I* parameter, $r = -.24$, 95% CI [−.35, −.13].

Discussion

Supporting Hypotheses 2a and 2b and consistent with the results of Study 1, participants who were more sensitive to moral norms were less likely to behave dishonestly by inflating their accuracy scores even when doing so would increase the amounts donated to a charity of their choice. Contrary to Hypotheses 3a and 3b, sensitivity to consequences was not significantly associated with dishonest behavior for the greater good. If anything, their association was in a direction that was opposite to the predicted positive association. Because sensitivity to consequence showed the predicted negative association with primary psychopathy, the obtained relation with dishonest behavior for the greater good cannot be attributed to psychometric issues associated with the *C* parameter. Exploratory analyses revealed no significant associations between dishonest behavior for the greater good and general action tendencies, as well as between dishonest behavior for the greater good and preference for utilitarian over deontological judgments as captured by the traditional dilemma score.

General discussion

The question of whether sacrificial dilemma judgments predict moral behavior remains contested in moral psychology. Past research contributing to the debate has typically assessed whether sacrificial dilemma judgments are predictive of morally relevant behavior captured by paradigms mirroring the classic trolley problem (e.g. Bostyn et al., 2018; Plunkett & Greene, 2019) or economic games (e.g. Bostyn & Roets, 2017; Capraro et al., 2018). Such research ensures a high correspondence between the assessed moral judgment and behavior, lending insight into the question of whether people do as they say. However, the results of this work may be distorted by measurement error given the reliance on single observations of moral behavior (e.g. Bostyn & Roets, 2017; Bostyn et al., 2018; Capraro et al., 2018). Moreover, tasks modeled after the classic trolley problem are known to conflate moral codes with action codes (Crone & Laham, 2017) and mistakenly

pit utilitarian judgments at odds with deontological judgments (Conway & Gawronski, 2013). Overcoming these limitations and going beyond the question of whether people do as they say, the current research contributes to the debate by examining if a broader moral disposition underlies both sacrificial dilemma judgments and other instances of morally relevant behavior. Toward this end, the current research used the CNI model of moral decision-making to investigate whether specific aspects of sacrificial dilemma judgments predict dishonest behavior enacted for personal gain (Study 1) and for the greater good (Study 2) over a series of trials.

Based on the assumptions that (1) the CNI model's *N* parameter may capture a broader disposition to conform to relevant moral norms in situations beyond sacrificial dilemmas and (2) lying to enhance task performance constitutes a violation of moral norms surrounding honesty, we expected to find a positive association between norm sensitivity and dishonest behavior in the coin-toss prediction task in both studies. Consistent with our predictions, sensitivity to moral norms in responses to sacrificial dilemmas was predictive of dishonest behavior in the coin-toss prediction task, regardless of whether the dishonesty led to potential personal gains (Study 1) or benefitted the greater good (Study 2). Specifically, individuals exhibiting stronger sensitivity to moral norms were less likely to engage in dishonest behavior by inflating their accuracy in predicting the outcomes of the coin tosses, and this relationship held even after we accounted for their sensitivity to consequences and general action tendencies. Notably, whereas scores on the HEXACO-60 honesty–humility scale were negatively correlated with dishonest behavior committed for personal gain (Study 1), honesty–humility was not significantly associated with dishonest behavior enacted for the greater good (Study 2). This finding suggests that the personality measure of honesty–humility is not a consistent, context-independent predictor of dishonest behavior. Instead, sensitivity to moral norms was the only factor consistently predicting whether one would behave dishonestly on the coin-toss prediction task.

Based on the assumptions that (1) the CNI model's *C* parameter may capture a broader disposition to maximize overall outcomes for the greater good in situations beyond sacrificial dilemmas and (2) inflating one's accuracy on the coin-toss prediction task increases benefits for the greater good, we expected to find a positive association between sensitivity to consequences and dishonest behavior in Study 2. Although our positive control measure provided empirical support for the psychometric quality of the *C* parameter, no support was found for the hypothesized relation between sensitivity to consequences and dishonest behavior for the greater good. This null finding suggests that sensitivity to consequences in responses to sacrificial dilemmas may not reflect a broader disposition to maximize overall outcomes for the greater good beyond sacrificial dilemmas. At the very least, sensitivity to consequences may reflect a utilitarian behavioral tendency that is distinct from the tendency to engage in concrete prosocial behaviors for the greater good. This conclusion is consistent with findings by Kahane et al. (2015), suggesting that utilitarian judgments in sacrificial dilemmas are weakly associated with utilitarian decisions in non-sacrificial contexts (e.g. donations). To account for their findings, Kahane et al. (2018) proposed a model of utilitarian moral decision-making that comprises two distinct dimensions: (1) instrumental harm, which reflects one's willingness to harm others for the greater good, and (2) impartial beneficence, which reflects one's impartial concern for the greater good. Research drawing on this distinction suggests that individual differences in sacrificial dilemma judgments primarily reflect differences along the instrumental-harm dimension

but not differences along the impartial-beneficence dimension (e.g. Kahane et al., 2018; but see, Körner et al., 2020). The null relation between sensitivity to consequences and dishonest behavior for the greater good could have been driven by this disparity, in that our sacrificial dilemmas had solely tapped the instrumental-harm dimension of utilitarian thinking, whereas our behavioral measure of dishonesty had solely tapped the impartial-beneficence dimension of utilitarian thinking. Thus, a potential conclusion from the current findings is that instrumental harm as captured by sacrificial dilemma judgments is not predictive of impartial beneficence in dishonest behavior for the greater good.

In the two studies, we did not make any specific predictions regarding the relations between dishonest behavior and (1) general action tendencies as captured by the CNI model's *I* parameter and (2) the traditional dilemma score, which considers responses to dilemmas wherein an action would violate a proscriptive moral norm but bring about greater benefits than costs (i.e. relative preference for utilitarian over deontological judgments). Our preregistered exploratory analyses revealed no associations between general action tendencies and dishonest behavior in both studies. Likewise, responses to the traditional dilemmas were not associated with dishonest behavior in the coin-toss prediction task. The null relation between the traditional dilemma responses and dishonest behavior highlights the value of the CNI model of moral decision-making in providing insights that cannot be gained with the traditional dilemma approach. By conflating moral codes with action codes, and by pitting deontology against utilitarianism, the traditional paradigm conflates different factors underlying sacrificial dilemma judgments. Because effects obtained with traditional dilemma scores are the combinatory result of multiple distinct determinants of sacrificial dilemma judgment, research using the traditional paradigm can lead to misleading conclusions, such as that responses to sacrificial dilemma judgments do not predict dishonest behavior. Only by using the CNI model to disentangle the different factors underlying sacrificial dilemma judgments were we able to identify the unique relationship between sensitivity to moral norms and dishonest behavior.

Our main finding that sensitivity to moral norms in responses to sacrificial dilemmas consistently predicted dishonest behavior suggests the operation of a broader underlying moral disposition guiding both sacrificial dilemma judgments and dishonest behavior. This finding has implications for the ongoing debate between pluralist and monist views in moral psychology. In the nomenclature of Moral Foundations Theory (MFT), our measure of sacrificial dilemma judgments pertained specifically to the moral foundation of harm, whereas our behavioral measure of dishonesty pertained to the moral foundation of fairness (Haidt & Graham, 2007). Because MFT claims that distinct mechanisms underlie different moral foundations (Graham et al., 2013), our finding that the factors underlying judgments of harm are predictive of moral behavior pertaining to fairness is not a trivial one from the perspective of MFT. Indeed, the results of the current research better align with the propositions of Gray et al.'s (2012) Theory of Dyadic Morality (TDM), which posits that the moral foundations identified by MFT are merely different content representations of harm (Schein & Gray, 2018). From the perspective of TDM, sacrificial dilemma judgments concerning harm should be predictive of moral behavior in other seemingly distal moral domains such as fairness, just as the current research has illustrated.

Qualifying the findings

An important caveat is that the identified relationships between the factors underlying sacrificial dilemma judgments and dishonest behavior may not generalize to all instances of morally relevant behavior. As aforementioned, different facets of utilitarian thinking can relate to various types of moral behaviors in distinct ways (Kahane et al., 2018, 2015). While sensitivity to consequences may not be associated with dishonest behavior enacted for charitable donations, it does not preclude an association with other types of morally relevant behavior that may involve instrumental harm. Future work exploring whether sensitivity to consequences predicts moral behavior more closely related to instrumental harm as opposed to impartial beneficence is thus needed. Likewise, research investigating whether the effects of sensitivity to moral norms generalize to a broad range of morally relevant behaviors would be insightful, as it would hint at the operation of an overarching moral disposition guiding both sacrificial dilemma judgments and the studied moral behaviors.

Another important caveat pertains to a recent critique of the CNI model by Baron and Goodwin (2020). Although some of Baron and Goodwin's concerns are based on misunderstandings of the model and flawed re-analyses of existing data (see, Gawronski et al., 2020), we discuss two important points and their implications for the current findings. First, it should be noted that the CNI model's *N* parameter captures norm-conforming judgments (i.e. judgments congruent with a given norm that do not necessarily result from explicit use of the norm), not norm-following judgments (i.e. judgments resulting from explicit use of a focal norm). As explained in more detail by Gawronski et al. (2020), the *N* parameter of the CNI model reflects the behavioral tendency to make judgments that conform to relevant moral norms when responding to sacrificial dilemmas, but this behavioral tendency may not be driven by explicit consideration of these norms. Moral norms in the CNI model dilemma battery are manipulated via pairs of descriptions of actions and inactions that either cause harm (i.e. presence of a proscriptive norm) or prevent harm (i.e. presence of a prescriptive norm), but that would otherwise result in the same outcomes. Conceptually, the *N* parameter simply reflects the extent to which participants' responses are influenced by this difference; the model does not assume that the responses captured by the *N* parameter are driven by conscious thoughts about specific moral norms.

Second, as acknowledged by Gawronski et al. (2017) and discussed in more detail by Baron and Goodwin (2020), the *I* parameter of the CNI model can be interpreted as an instance of deontological responding in the sense that it captures inaction responses conforming to the broad principle *first, do no harm*. Thus, it is important to emphasize that the relationship between the *N* parameter and dishonest behavior obtained in the current studies specifically reflects the effects of one's behavioral tendency to adhere to proscriptive and prescriptive moral norms as specified above. It does not reflect one's tendency to adhere to the broad principle *first, do no harm*, which should have been reflected in a significant association between dishonest behavior and the *I* parameter.

Conclusion

The current research addressed three methodological issues of extant work investigating the predictive utility of sacrificial dilemma judgments. Our findings revealed a consistent relationship between sensitivity to moral norms and dishonest behavior, suggesting the operation of a broader underlying moral disposition. By identifying such distal associations between specific aspects of sacrificial dilemma judgments and dishonest behavior enacted under different contextual reward conditions, the current research demonstrates the value of sacrificial dilemma research for understanding instances of morally relevant behavior and the more nuanced insights that can be gained from using the CNI model to disentangle different aspects of sacrificial dilemma judgments.

Notes

1. In actuality, all participants received the larger sum of money in compensation for their participation.
2. Some studies have compared moral dilemma judgments on standard text-based formats to behaviors in corresponding dilemmas simulated using virtual reality (VR) technology (e.g. Francis et al., 2016; Patil et al., 2014). However, this work has focused mainly on mean-level differences between judgment and behavior rather than associations between judgment and behavior.
3. The reported average correlations were based on a meta-analysis of the six studies from our group using Levenson et al.'s (1995) PPS (Gawronski et al., 2017; Körner et al., 2020; Luke & Gawronski, 2021b). We did not have any unpublished data from studies using the PPS at the time we conducted the current research.
4. Power analyses were conducted using G*Power 3.1.9.2 (Faul et al., 2007).
5. To avoid artifacts from lack of realism and plausibility (see, Kneer & Hannikainen, 2022; Körner et al., 2019), the CNI model dilemmas have been designed to be more realistic and plausible compared to the trolley problem and its variants (see, Gawronski et al., 2017; Körner et al., 2020). Nevertheless, a considerable number of studies using the CNI model dilemmas has found effects on preference for utilitarian over deontological judgments that replicate earlier findings using the trolley paradigm, including (but not limited to) effects of cognitive resources (Gawronski et al., 2017), incidental happiness (Gawronski et al., 2018), empathic concern (Körner et al., 2020), psychopathy (Luke & Gawronski, 2021b), and political ideology (Luke & Gawronski, 2021a).
6. 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 studies.
7. Participants were also told that, in the case of ties, we would randomly choose among tied participants to receive the $50 bonus payment.
8. Because $p(\text{action}) = 1 - p(\text{inaction})$, there are only four non-redundant equations in the full set of eight equations in the Appendix.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Nyx L. Ng http://orcid.org/0000-0002-3416-1385
Dillon M. Luke http://orcid.org/0000-0001-8784-3251
Bertram Gawronski http://orcid.org/0000-0001-7938-3339

References

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

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

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

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. https://doi.org/10.1111/spc3.12131

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. doi:10.1037/a0021524.

Bostyn, D. H., & Roets, A. (2017). Trust, trolleys and social dilemmas: A replication study. *Journal of Experimental Psychology. General*, *146*(5), e1–e7. https://doi.org/10.1037/xge0000295

Bostyn, D. H., & Roets, A. (2019). Should trolleys be scared of mice? Replies to Evans and Brandt (2019); Białek, Turpin, and Fugelsang (2019); Colman, Gold, and Pulford (2019); and Plunkett and Greene (2019). *Psychological Science*, *30*(9), 1392–1396. https://doi.org/10.1177/0956797619865236

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, *29*(7), 1084–1093. https://doi.org/10.1177/0956797617752640

Bostyn, D. H., Sevenhant, S., & Roets, A. (2019). Beyond physical harm: How preference for consequentialism and primary psychopathy relate to decisions on a monetary trolley dilemma. *Thinking & Reasoning*, *25*(2), 192–206. https://doi.org/10.1080/13546783.2018.1497536

Canty, A., & Ripley, B. D. (2021). *Boot: bootstrap R (S-Plus) functions*. R package.

Capraro, V., Sippel, J., Zhao, B., Hornischer, L., Savary, M., Terzopoulou, Z., Faucher, P., & Griffioen, S. F. (2018). People making deontological judgments in the Trapdoor dilemma are perceived to be more prosocial in economic games than they actually are. *PLOS ONE*, *13*(10), e0205066. doi:10.1371/journal.pone.0205066.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum. https://doi.org/10.4324/9780203771587

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

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. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366. https://doi.org/10.1016/j.tics.2013.06.005

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

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

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269. https://doi.org/10.1016/j.tics.2020.01.007

Dickinson, D. L., & Masclet, D. (2019). Using ethical dilemmas to predict antisocial choices with real payoff consequences: An experimental study. *Journal of Economic Behavior & Organization*, *166*, 195–215. https://doi.org/10.1016/j.jebo.2019.08.023

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

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

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLOS ONE*, *11*(10), e0164374. doi:10.1371/journal.pone.0164374.

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., Armstrong, J., Friesdorf, R., & Hütter, M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion*, *18*(7), 989–1008. https://doi.org/10.1037/emo0000399

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.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In. Devine, Patricia, Plant, Ashby (eds.). *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, *23*(2), 206–215. https://doi.org/10.1080/1047840X.2012.686247

Greene, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 3, pp. 35–80). MIT Press. https://doi.org/10.7551/mitpress/7504.003.0004

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences*, *106*(30), 12506–12511. https://doi.org/10.1073/pnas.0900152106

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), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, *13*(4), 356–371.

Hester, N., & Gray, K. (2020). The moral psychology of raceless, genderless strangers. *Perspectives on Psychological Science*, *15*(2), 216–230. https://doi.org/10.1177/1745691619885840

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*, 72–88. https://doi.org/10.1016/j.jrp.2015.04.003

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340–1343. https://doi.org/10.1126/science.1251560

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. doi:10.1037/rev0000093.

Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, *134*, 193–209. https://doi.org/10.1016/j.cognition.2014.10.005

Kneer, M., & Hannikainen, I. R. (2022). Trolleys, triage and Covid-19: The role of psychological realism in sacrificial dilemmas. *Cognition & Emotion*, *36*(1), 137–153. https://doi.org/10.1080/02699931.2021.1964940

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911. https://doi.org/10.1038/nature05631

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

Körner, A., Joffe, S., & Deutsch, R. (2019). When skeptical, stick with the norm: Low dilemma plausibility increases deontological moral judgments. *Journal of Experimental Social Psychology*, *84*, 103834. https://doi.org/10.1016/j.jesp.2019.103834

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://psycnet.apa.org/doi/10.1037/0022-3514.68.1.151

Luke, D. M., & Gawronski, B. (2021a). Political ideology and moral dilemma judgments: An analysis using the CNI model. *Personality & 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. (2021c). Temporal stability of moral dilemma judgments: A longitudinal analysis using the CNI model. *Personality & Social Psychology Bulletin 48*(8), 1191–1203. https://doi.org/10.1177/01461672211035024

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. doi:10.1037/per0000226.

Moore, A. B., Stevens, J., & Conway, A. R. A. (2011). Individual differences in sensitivity to reward and punishment predict moral judgment. *Personality and Individual Differences*, *50*(5), 621–625. https://doi.org/10.1016/j.paid.2010.12.006

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

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

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

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

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. doi:10.1016/j.jesp.2017.01.006.

Plunkett, D., & Greene, J. D. (2019). Overlooked evidence and a misunderstanding of what trolley dilemmas do best: Commentary on Bostyn, Sevenhant, and Roets (2018). *Psychological Science*, *30*(9), 1389–1391. https://doi.org/10.1177/0956797619827914

Schein, C. (2020). The importance of context in moral judgments. *Perspectives on Psychological Science*, *15*(2), 207–215. https://doi.org/10.1177/1745691620904083

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–70. https://doi.org/10.1177/1088868317698288

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Shalvi, S., & De Dreu, C. K. W. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences*, *111*(15), 5503–5507. https://doi.org/10.1073/pnas.1400724111

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

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415. https://doi.org/10.2307/796133

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(6), 476–477. https://doi.org/10.1111/j.1467-9280.2006.01731.x

Appendix: CNI Model Equations

Model equations for the estimation of sensitivity to consequences (*C*), sensitivity to moral norms (*N*), and general preference for inaction versus action (*I*) in responses to moral dilemmas with proscriptive versus prescriptive norms and benefits of action for overall well-being that are either greater or smaller than the costs of action for well-being. Reproduced from Gawronski et al. (2017). Reprinted with permission from the American Psychological Association.

$$p(\textit{inaction}|\ \textit{proscriptive norm},\ \textit{benefits}\ >\ \textit{costs}) = [(1-C)\times N] + [(1-C)\times(1-N)\times I]$$

$$p(\textit{inaction}|\ \textit{proscriptive norm},\ \textit{benefits} < \textit{costs}) = C + [(1-C)\times N] + [(1-C)\times(1-N)\times I]$$

$$p(\textit{inaction}|\textit{proscriptive norm},\ \textit{benefits} > \textit{costs}) = (1-C)\times(1-N)\times I$$

$$p(\textit{inaction}|\ \textit{proscriptive}|\textit{norm},\ \textit{benefits} < \textit{costs}) = C + [(1-C)\times(1-N)\times I]$$

$$p(\textit{action}|\ \textit{proscriptive norm},\ \textit{benefits} > \textit{costs}) = C + [(1-C)\times(1-N)\times(1-I)]$$

$$p(\textit{action}|\ \textit{proscriptive norm},\ \textit{benefits} < \textit{costs}) = (1-C)\times(1-N)\times(1-I)$$

$$p(\textit{action}|\ \textit{proscriptive norm},\ \textit{benefits}\ >\ \textit{costs}) = C + [(1-C)\times N] + [(1-C)\times(1-N)\times(1-I)]$$

$$p(\textit{action}|\ \textit{proscriptive norm},\ \textit{benefits}\ <\textit{costs}) = [(1-C)\times N] + [(1-C)\times(1-N)\times(1-I)]$$