Judgment and Decision Making, Vol. 15, No. 6, November 2020, pp. 1054–1072

On the validity of the CNI model of moral decision-making: Reply to Baron and Goodwin (2020)

Bertram Gawronski* Paul Conway† Mandy Hütter‡ Dillon M. Luke§ Joel Armstrong¶ Rebecca Friesdorf‖

Abstract

The CNI model of moral decision-making is a formal model that quantifies (1) sensitivity to consequences, (2) sensitivity to moral norms, and (3) general preference for inaction versus action in responses to moral dilemmas. Based on a critique of the CNI model's conceptual assumptions, properties of the moral dilemmas for research using the CNI model, and the robustness of findings obtained with the CNI model against changes in model specifications, Baron and Goodwin (2020) dismissed the CNI model as a valid approach to study moral dilemma judgments. Here, we respond to their critique, showing that Baron and Goodwin's dismissal of the CNI model is based on: (1) misunderstandings of key aspects of the model; (2) a conceptually problematic conflation of behavioral effects and explanatory mental constructs; (3) arguments that are inconsistent with empirical evidence; and (4) reanalyses that supposedly show inconsistent findings resulting from changes in model specifications, although the reported reanalyses did not actually use the CNI model and proper analyses with the CNI model yield consistent findings across model specifications. Although Baron and Goodwin's critique reveals a need for greater precision in the description of the three model parameters and for greater attention to properties of individual dilemmas, the available evidence indicates that the CNI model is a valid, robust, and empirically sound approach to gaining deeper insights into the determinants of moral dilemma judgments, overcoming major limitations of the traditional approach that pits moral norms against consequences for the greater good (e.g., trolley dilemma).

Keywords: CNI model; deontology; moral dilemmas; moral judgment; multinomial modeling; omission bias; utilitarianism

1 Introduction

A central question in moral psychology is how people make decisions in moral dilemmas that involve a conflict between moral norms and the greater good. Research on this question has predominantly relied on hypothetical scenarios such as the trolley problem, in which a runaway trolley is approaching a group of five individuals who would be killed if the trolley continues on its path. In one variant known as the switch dilemma, participants are asked if it would be acceptable to pull a switch to redirect the trolley to another track where it would kill only one person instead of five (Foot, 1967). In another variant known as the footbridge dilemma, participants are asked if it would be acceptable to push a person from a bridge to stop the trolley (Thomson, 1971). Adopting terminology from moral philosophy, participants are said to have made a characteristically utilitarian judgment if they judge the described actions as acceptable. Conversely, participants are said to have made a characteristically deontological judgment if they judge the described actions as unacceptable (Conway, Goldstein-Greenwood, Polacek & Greene, 2018). From a utilitarian view, the described actions would be acceptable because they maximize the well-being of a larger number of people. In contrast, from a deontological view, the described actions would be unacceptable because they are in conflict with the moral norm that one should not kill other people. Over the past two decades, a substantial amount of research has investigated contextual conditions that make people more or less likely to prefer utilitarian over deontological judgments, individual differences in the preference for utilitarian over deontological judgments, and the mental processes underlying utilitarian and deontological judgments (Bartels, Bauman, Cushman, Pizarro & McGraw, 2015).

Research using the trolley problem (and similar sacrificial dilemmas) has been criticized for relying on unrealistic, sometimes humorous scenarios that have little resemblance with the kinds of moral dilemmas people are facing

This research was supported by National Science Foundation (NSF) Grant BCS-1449620 to Bertram Gawronski and a Heisenberg grant (HU 1978/7-1) from the German Research Foundation to Mandy Hütter. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the funding agencies.

*University of Texas at Austin. Email: gawronski@utexas.edu.
†Florida State University .
‡Eberhard Karls Universität Tübingen.
§University of Texas at Austin.
¶University of Western Ontario.
‖Wilfrid Laurier University.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

in real-world contexts (Bauman, McGraw, Bartels & Warren, 2014; Körner, Joffe & Deutsch, 2019). In addition to addressing this concern by using realistic dilemmas inspired by real-world cases, our own research aimed to resolve a more fundamental limitation of the traditional dilemma approach: the confounding of multiple factors in the measurement of moral judgments. To resolve these confounds, we have developed a mathematical model called the CNI model of moral decision-making, which quantifies three determinants of moral dilemma judgments: (1) sensitivity to consequences, (2) sensitivity to moral norms, and (3) general preference for inaction versus action (Gawronski, Armstrong, Conway, Friesdorf & Hütter, 2017). Based on a critique of this work, Baron and Goodwin (2020) dismissed the CNI model as a valid approach to studying moral dilemma judgments. Here, we respond to their criticism, clarifying which of their concerns are justified and which ones are not. Although Baron and Goodwin's critique reveals a need for greater precision in the description of the three model parameters and for greater attention to properties of individual dilemmas, we argue that Baron and Goodwin's dismissal of the CNI model is based on: (1) misunderstandings of key aspects of the model; (2) a conceptually problematic conflation of behavioral effects and explanatory mental constructs; (3) arguments that are inconsistent with empirical evidence; and (4) reanalyses that supposedly show inconsistent findings resulting from changes in model specifications, although the reported reanalyses did not actually use the CNI model and proper analyses with the CNI model yield consistent findings across model specifications. We conclude that, counter to Baron and Goodwin's conclusion, the CNI model is a valid, robust, and empirically sound approach to gaining deeper insights into the determinants of moral dilemma judgments, overcoming major limitations of the traditional approach.

2 The CNI Model

Because Baron and Goodwin's (2020) critique is partly based on misunderstandings of key aspects of the CNI model, we deem it important to provide some background information about the model for our rebuttal of their critique. The CNI model is based on the notion that utilitarian judgments are characterized by the feature of being influenced by the consequences of a given action for the greater good, whereas deontological judgments are characterized by the feature of being influenced by the consistency of a given action with moral norms (Gawronski & Beer, 2017). Thus, to classify moral judgments as utilitarian, one would need to demonstrate that they vary as a function of relevant consequences for the greater good. Conversely, to classify moral judgments as deontological, one would need to demonstrate that they vary as a function of relevant moral norms. Whereas the former requires experimental manipulations of consequences, the latter requires experimental manipulations of moral norms. Both are largely absent in research using the traditional dilemma approach, which focuses almost exclusively on (1) cases where the benefits of a focal action for the greater good outweigh its costs (without considering cases where the benefits of the focal action are smaller than its costs) and (2) actions that are prohibited by proscriptive norms (without considering actions that are prescribed by prescriptive norms).

These limitations pose a significant challenge to the interpretation of moral dilemma judgments. For example, if a participant finds it acceptable to redirect the trolley in the switch dilemma regardless of whether it would save five lives or only one, questions could be raised about whether it is justified to categorize the participant's responses as utilitarian, because they are unaffected by the consequences for the greater good. Similarly, if a participant is unwilling to perform a focal action regardless of whether this action is prohibited by a proscriptive norm (e.g., killing a person) or prescribed by a prescriptive norm (e.g., saving a person's life), questions could be raised about the deontological nature of these judgments, because they may reflect general action aversion rather than effects of moral norms.

A different way of describing these issues is that the traditional approach includes two confounds in the measurement of moral dilemma judgments (Gawronski, Conway, Armstrong, Friesdorf & Hütter, 2016). First, because accepting one option implies rejecting the other, it is impossible to determine whether differences in moral dilemma judgments are driven by differences in the tendency to make a "utilitarian" judgment, differences in the tendency to make a "deontological" judgment, or differences in both (Conway & Gawronski, 2013). Second, because research using the traditional dilemma approach has focused almost exclusively on cases involving proscriptive norms, "utilitarian" judgments are confounded with general preference for action and "deontological" judgments are confounded with general preference for inaction (Crone & Laham, 2017).

To resolve these interpretational ambiguities, we proposed an alternative approach in which responses are compared across four types of dilemmas that vary in terms of whether (1) the consequences of the focal action for the greater good are either greater or smaller than the costs and (2) the focal action is either proscribed by a proscriptive norm or prescribed by a prescriptive norm. Expanding on this approach, we developed a multinomial model that quantifies the extent to which participants' responses across the four types of dilemmas reflect a response pattern that is sensitive to consequences (see first row in Figure 1), a response pattern that is sensitive to moral norms (see second row in Figure 1), and a general preference for inaction versus action (see third and fourth row in Figure 1). Sensitivity to consequences is captured by the model's *C* parameter, with higher scores reflecting a greater impact of consequences on moral judg-

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

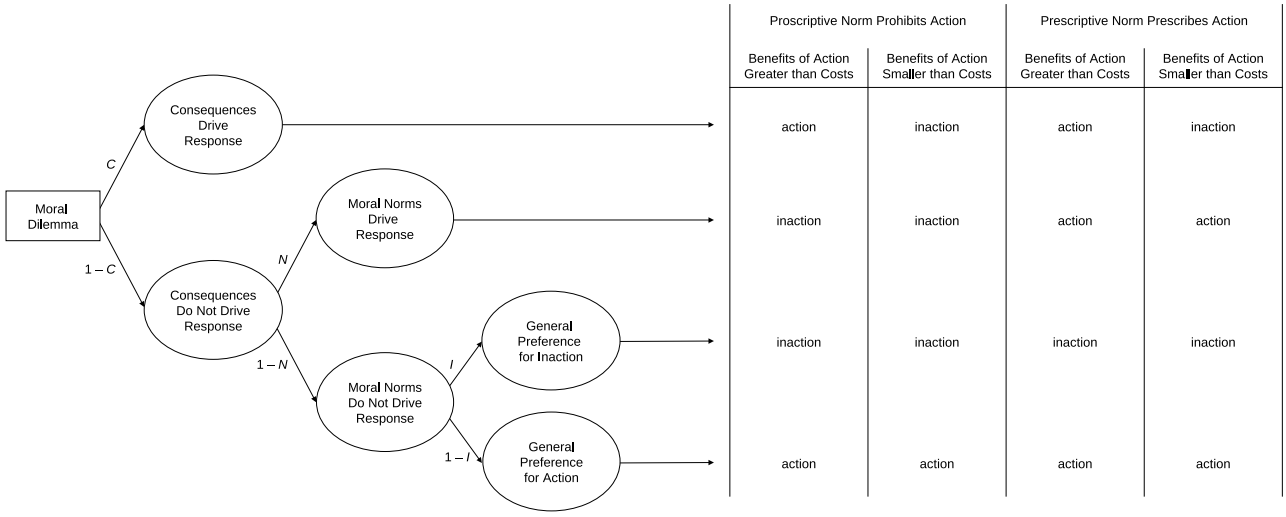

FIGURE 1: CNI model of moral decision-making predicting action versus inaction responses in moral dilemmas with proscriptive and prescriptive norms and consequences involving benefits of action that are either greater or smaller than the costs of action. Reproduced from Gawronski, Armstrong, Conway,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

ments. Sensitivity to moral norms is captured by model's *N* parameter, with higher scores reflecting a greater impact of moral norms on moral judgments. General preference for inaction versus action is captured by the model's *I* parameter, with higher scores reflecting a general preference for inaction and lower scores reflecting a general preference for action.

Based on the processing tree depicted in Figure 1, the CNI model provides four mathematical equations that include the three model parameters as unknowns and the observed probabilities of action (vs. inaction) responses on the four kinds of dilemmas as known numerical values (see Appendix A).[1] Numerical values for the three parameters are estimated via maximum likelihood statistics, aiming to minimize the difference between the empirically observed probabilities of action (vs. inaction) responses on the four types of dilemmas and the probabilities of action (vs. inaction) responses predicted by the model equations using the identified parameter estimates. The estimated scores for each parameter reflect a probability that can vary between 0 and 1. For the *C* parameter, scores significantly greater than zero indicate that responses were affected by the manipulation of consequences in a manner such that participants showed a response pattern that maximizes the greater good. For the *N* parameter, scores significantly greater than zero indicate that responses were affected by the manipulation of moral norms such that participants showed a response pattern that is congruent with both proscriptive and prescriptive norms. Finally, for the *I* parameter, scores significantly greater than 0.5 indicate a general preference for inaction and scores significantly lower than 0.5 indicate a general preference for action. The adequacy of the model in describing the data can be evaluated by means of goodness-of-fit statistics, such that poor model fit would be reflected in a statistically significant deviation between the empirically observed probabilities and the probabilities predicted by the model (for more details, see Gawronski et al., 2017).

[1]Note that there are only four non-redundant equations among the eight equations in the Appendix, because $p(\text{action}) = 1 - p(\text{inaction})$.

3 Deontological Responding

A central point of Baron and Goodwin's (2020) critique is that bias against action, as captured by the *I* parameter, is an explanation of deontological responding rather than an alternative process. Although we would argue that bias against action is not a mental process but a pattern of responding, we fully agree that a general preference for inaction can be interpreted as a particular instance of deontological responding. In fact, we already acknowledged such an interpretation in the original publication on the CNI model when we discussed the (in)consistency of our findings with Greene's (2008) dual-process theory of moral dilemma judgments (Gawronski et al., 2017, p. 365):

"A potential way to reconcile [our findings] with the dual-process model is to interpret general preference for inaction as an instance of deontological responding. In line with this idea, the doctrine of doing and allowing (DDA) states that actively causing harm is morally worse than merely allowing harm, which is consistent with the finding that harm caused by action is perceived as worse than equivalent harm caused by inaction [. . .]. Conceptually, the DDA can be regarded as

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

a deontological principle in the sense that the moral status of a behavioral option depends on its consistency with a general rule."

Baron and Goodwin's explication of this idea is much more detailed compared to our analysis, and we agree with almost every aspect of their arguments about a deontological interpretation of generalized inaction. That being said, we would like to note that a conceptualization of generalized inaction as an instance of deontological responding remains incomplete if it does not acknowledge adherence to both proscriptive and prescriptive norms as a distinct pattern of deontological responding over and above generalized inaction. Although both response patterns can be interpreted as instances of deontological responding in a philosophical sense, they are conceptually distinct in the sense that they involve different behavioral signatures. As we stated in the original publication on the CNI model (Gawronski et al., 2017, p. 365):

"Although the two ways of responding may be deemed deontological in a philosophical sense, they should not be conflated in a psychological theory about the mechanisms underlying moral dilemma judgment. After all, sensitivity to moral norms and general preference for inaction are functionally distinct in terms of their psychological antecedents and their behavioral outcomes. Their distinct outcomes are reflected in the fact that the two ways of responding lead to different judgments in moral dilemmas involving a prescriptive norm. Their distinct antecedents are reflected in the current finding that a given factor can simultaneously strengthen one way of 'deontological' responding while weakening the other way of 'deontological' responding."

Thus, although we agree with Baron and Goodwin's (2020) argument that bias against action can be interpreted as an instance of deontological responding, their dominant focus on generalized inaction ignores that adherence to proscriptive and prescriptive norms represents a distinct pattern of deontological responding that does not involve generalized inaction. That being said, there is a nuance to Baron and Goodwin's argument that suggests potentially misleading aspects in the descriptions we have chosen for the response patterns captured by the *N* and the *I* parameter.

By describing the response pattern captured by the *N* parameter as *sensitivity to moral norms* and the response pattern captured by the *I* parameter as *general preference for inaction versus action*, our descriptions suggest that only the former, but not the latter, response pattern would be congruent with moral norms. However, as correctly noted by Baron and Goodwin (2020), a pattern of generalized inaction is congruent with the broad deontological norm *first, do no harm,* suggesting that congruence with moral norms is not a distinguishing feature of the two parameters. Baron and Goodwin (2020) are also correct in their observation that the response pattern captured by the *I* parameter may reflect either a domain-specific aversion against causing harm or a domain-independent response bias (see Hennig & Hütter, 2020). Based on these considerations, we acknowledge that some aspects of our parameter descriptions are not ideal and potentially misleading, which includes some of the descriptions in the processing tree depicted in Figure 1 (e.g., the description of $1 - N$ as *moral norms do not drive response*). For the sake of consistency with prior CNI model terminology, we will continue to describe the response pattern captured by the *N* parameter as *sensitivity to moral norms*, yet with the qualification that this description is meant to refer specifically to the effect of proscriptive versus prescriptive norms on moral judgments (rather than a pattern of generalized inaction that is congruent with the broad norm *first, do no harm*). Moreover, the description *general preference for inaction versus action* should be interpreted in a purely behavioral manner that is agnostic about whether generalized inaction on the *I* parameter reflects a domain-specific aversion against causing harm or a domain-independent response bias. These qualifications are important for the interpretation of the two parameters, because they imply that the *I* parameter may capture a norm-congruent response pattern that is distinct from the norm-congruent response pattern captured by the *N* parameter. Yet, regardless of these qualifications, we would still argue that both response patterns are essential for understanding responses to moral dilemmas and their underlying mental processes. We will return to this issue in the section entitled "Confounds in the Traditional Approach."

4 Moral Norms

A bias against action is relatively easy to identify, because it involves a general preference for inaction regardless of the specific situation (see third row in Figure 1). In contrast, differential responses to dilemmas with proscriptive and prescriptive norms are more difficult to identify, because any such endeavor requires construct-valid operationalizations of the two kinds of moral norms (see second row in Figure 1). A second major point of Baron and Goodwin's (2020) critique is that our battery of moral dilemmas for research using the CNI model does not meet this criterion.

To address this concern, we deem it helpful to first clarify how we identified the focal actions and their corresponding moral norms in the construction of CNI model dilemmas. In a first step, we identified pairs of morally relevant actions and inactions that have the same outcome (e.g., killing Person A and letting Person A die both result in the loss of Person A's life). Whereas the identified action within each action-inaction pair was conceptually linked to a proscriptive norm (i.e., killing someone is morally prohibited), the opposite of the identified inaction was conceptually linked to a prescriptive norm (i.e., saving someone's live is morally prescribed). In a second step, we generated hypothetical consequences of the two actions that involve costs for the well-being of others

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

that are either greater or smaller than the benefits of each action. In a third step, we created plausible scenarios of high real-world relevance, which were designed to be as similar as possible in the four variants of each basic dilemma. Thus, based on the rationale underlying the development of CNI model dilemmas, the relevant moral norms and their corresponding actions are identified in the first step independent of their consequences and independent of secondary aspects of the particular scenarios. In all cases, the relevant proscriptive norms pertain to directly causing harm regardless of the situation (e.g., killing someone), whereas the relevant prescriptive norms pertain to directly preventing harm regardless of the situation (e.g., saving someone's life).

Baron and Goodwin (2020) correctly point out that prescriptive norms tend to be weaker than proscriptive norms (e.g., killing someone is perceived as morally worse than letting someone die). Baron and Goodwin conclude from this asymmetry that our modeling approach is destined to fail, because it requires a switch of actions and inactions while maintaining equivalent norms, which seems virtually impossible. Although we agree that the asymmetry between proscriptive and prescriptive norms is fundamentally important, it is irrelevant for the construction of CNI model dilemmas to the extent that the proximal outcomes of a given action-inaction pair can be held constant (e.g., loss of the same life as a result of killing or letting die). From the perspective of the CNI model, the possibility of symmetric and asymmetric effects of the two kinds of norms is not a methodological obstacle, but an empirical phenomenon that is captured by the difference between the *N* and the *I* parameter. Whereas the response pattern defining the *N* parameter is congruent with both proscriptive and prescriptive norms, the response pattern defining the *I* parameter is congruent with only one of the two norms but not the other (with the specific congruency depending on whether scores are greater or smaller than 0.5; see Figure 1).

A related concern raised by Baron and Goodwin (2020) is that a considerable number of participants seem to disagree with our assumptions about relevant moral norms in the CNI model dilemmas. To support their argument, Baron and Goodwin present the results of two studies in which participants were asked to identify for a selected subset of CNI model dilemmas (and a set of newly created dilemmas) if there is a rule that favors a particular response and, if so, which response is favored by that rule. Their main finding is that many participants identified patterns of rules that conflict with the conceptual assumptions underlying the operationalization of moral norms in our CNI model dilemmas, which led them to question the validity of our operationalization.

A major problem with Baron and Goodwin's (2020) argument and the presented data is that they conflate different levels of analysis. As we explained in Gawronski, Conway, Armstrong, Friesdorf and Hütter (2018), the CNI model is a descriptive model that quantifies patterns of stimulus-response relations at the behavioral level of analysis (see De Houwer, 2011; Gawronski & Bodenhausen, 2015). Applied to the operationalization of moral norms in our dilemmas, the "stimulus" involves descriptions of actions that cause harm (capturing the presence of a proscriptive norm) or descriptions of actions that prevent harm (capturing the presence of a prescriptive norm). The "response" involves participants' judgments of the described actions (e.g., acceptable or unacceptable). The *N* parameter merely captures the extent to which participants' responses differ across the two cases, which reflects their sensitivity to the two kinds of norms in responding to the moral dilemmas (see Figure 1). The CNI model does not make any assumptions about the mental processes underlying the observed response pattern. The latter question pertains to the mental level of analysis, which aims to identify the mental processes underlying observed patterns of stimulus-response relations at the behavioral level (see De Houwer, 2011; Gawronski & Bodenhausen, 2015).

Baron and Goodwin's (2020) critique is based on the tacit background assumption that an experimental manipulation of moral norms is construct-valid only if the behavioral effect of this manipulation is driven by conscious thoughts about moral norms. Their data suggest that this is not the case for our manipulation of moral norms, because it does not map onto the moral norms identified by the participants in their studies. However, by requiring that conscious thoughts about moral norms must underlie norm-congruent judgments, Baron and Goodwin's (2020) critique not only conflates behavioral effects with explanatory mental constructs (see De Houwer, 2011; Gawronski & Bodenhausen, 2015); it also ignores one of the most significant contributions to moral psychology in the 21st century: the idea that norm-congruent judgments may not necessarily be the product of conscious thoughts about norms (Greene, 2008; Haidt, 2001). For example, according to Greene's (2008) dual-process theory of moral dilemma judgments, norm-congruent judgments are driven by affective processes that do not involve conscious thoughts about moral norms. As we noted above, the CNI model is a descriptive model of stimulus-response relations at the behavioral level and the model does not make any assumptions about underlying processes at the mental level (see Gawronski et al., 2018). Hence, the pattern of norm-congruent responses captured by the *N* parameters could be driven by conscious thoughts about moral norms, affective responses to the idea of causing harm, or something entirely different. From this perspective, involvement of conscious thoughts about moral norms is not a suitable criterion to evaluate the validity of our experimental manipulation of moral norms, because people may behave in line with moral norms without consciously identifying their responses as reflecting moral norms.

If mental processes are inadequate to determine the validity of our norm manipulation, what is a good alternative to evaluate its validity? A relatively simple criterion is

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

TABLE 1: Goodness-of-fit statistics for the CNI model in all studies published by the current authors that have used the original battery of 24 moral dilemmas for research using the CNI model.

Reference	Study #	N	G^2	df	p
Bialek, Paruzel-Czachura & Gawronski (2019)	1	634	14.29	8	.074
Brannon, Carr, Jin, Josephs & Gawronski (2019)	1	200	0.79	2	.675
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	1a	201	1.32	2	.517
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	1b	197	1.51	2	.469
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	2a	194	4.98	2	.083
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	2b	194	1.29	2	.524
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	3a	186	11.93	2	.003
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	3b	189	4.19	2	.123
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	4a	184	0.29	2	.864
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	4b	198	0.18	2	.916
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	S1a	195	0.68	2	.713
Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017)	S1b	191	0.84	2	.656
Gawronski & Brannon (2020)	1a	140	2.72	2	.257
Gawronski & Brannon (2020)	1b	120	0.11	2	.945
Gawronski & Brannon (2020)	2a	91	2.52	2	.284
Gawronski & Brannon (2020)	2b	120	0.41	2	.817
Gawronski & Brannon (2020)	3	255	12.15	4	.016
Gawronski, Conway, Armstrong, Friesdorf & Hütter (2018)	1a	128	2.08	2	.354
Gawronski, Conway, Armstrong, Friesdorf & Hütter (2018)	1b	120	0.74	2	.691
Gawronski, Conway, Armstrong, Friesdorf & Hütter (2018)	2a	119	0.06	2	.972
Gawronski, Conway, Armstrong, Friesdorf & Hütter (2018)	2b	120	0.34	2	.842
Gawronski, Conway, Armstrong, Friesdorf & Hütter (2018)	3a	120	2.08	2	.353
Gawronski, Conway, Armstrong, Friesdorf & Hütter (2018)	3b	120	2.14	2	.343

that, although participants may not consciously think about moral norms in a manner that corresponds to our experimental manipulation, they should respond to the dilemmas in a manner "as if" they agree with the assumptions underlying our operationalization of moral norms (see above). In other words, participants should show evidence for *rule-conforming* judgments, even if they do not show any evidence for *rule-following* judgments. Multinomial modeling offers two pieces of evidence that are relevant for this question.

First, if participants did not respond in a manner "as if" they agreed with the assumptions underlying our operationalization of moral norms, the CNI model should be unable to provide accurate descriptions of the data, and thus show poor model fit across studies (Klauer, 2015). Counter to this possibility, the CNI model fit the data well in almost every published study that we conducted with our original dilemma battery (see Table 1). Out of 23 published studies, there are only two studies in which the response probabilities predicted by the CNI model significantly deviated from the response probabilities observed in the data. With an alpha-criterion of .05, this proportion is close to the probability of obtaining a significant difference in the absence of an actually existing difference. If we include every unpublished study from our group using the original CNI model battery, the proportion is even lower with only two out of 32 studies in which the model did not fit the data.

Second, if participants did not respond in a manner "as if" they agreed with the assumptions underlying our operationalization of moral norms, the N parameter should not significantly differ from zero (Klauer, 2015). Counter to this possibility, the N parameter was significantly greater than zero in every single study we have conducted with the CNI model. The only exception is a subsample of participants who were preselected based on having psychopathy scores that placed them in the highest quartile of participants in a broader sample (Gawronski et al., 2017, Study 4b). We will return to this finding in the section entitled "'Perverse' Responses." Together, the two pieces of evidence support the assumption that participants respond in a manner "as if" they agree with the assumptions underlying our operationalization of moral norms.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

5 Item Characteristics

The reported data on model fit and the significant difference of the *N* parameter from zero suggest that, on average, participants respond to the CNI model dilemmas in a manner that is consistent with the assumptions underlying our operationalization of moral norms. However, these results are based on aggregate responses to all dilemmas. Although aggregation of responses across multiple items has the advantage of reducing the likelihood of false negatives by controlling for measurement error, it can lead to false positives if an observed effect is driven by a subset of items with idiosyncratic features that are unrelated to the construct of interest (Judd, Westfall & Kenny, 2012). As noted by Baron and Goodwin (2020), multinomial modeling does not consider potential differences between individual dilemmas and their contribution to the obtained results. Although the problems associated with item-specific effects have received increased attention during the last years, there is no straightforward way to statistically control for item-specific effects within multinomial modeling. This limitation raises the question of whether the items of our dilemma battery are equally valid in capturing the manipulations of consequences and moral norms.

Baron and Goodwin (2020) aimed to address this question by analyzing item-specific proportions of "perverse" responses, which they define as responses that go against both moral norms and the greater good when the two suggest the same decision. However, from the perspective of the CNI model, this strategy remains uninformative about the validity of a given dilemma, because incidental features of a given scenario may shift responses toward action or inaction for all four variants without qualifying the effects of consequences and moral norms (see Schwarz, 1999). In other words, a given item may show "perverse" responses in terms of Baron and Goodwin's criterion, but the four variants of a given dilemma may still reliably capture the manipulations of consequences and moral norms. Thus, to determine whether the items of our dilemma battery are equally valid in capturing the manipulations of consequences and moral norms, it is essential to analyze whether the two manipulations are effective in influencing responses on each basic dilemma. For the manipulation of consequences, this question boils down to the proportion of action (vs. inaction) responses on dilemmas where the benefits of action are greater than the costs compared to the proportion of action (vs. inaction) responses on dilemmas where the benefits of action are smaller than the costs. For the manipulation of moral norms, the question boils down to the proportion of action (vs. inaction) responses on dilemmas where a proscriptive norm prohibits action compared to the proportion of action (vs. inaction) responses on dilemmas where a prescriptive norm prescribes action. The two comparisons should reveal meaningful differences in the expected direction for each basic dilemma across the four variants.

Using these criteria, our pilot tests supported the validity of each basic dilemma that was included in the final battery of dilemmas for research using the CNI model (some other dilemmas were discarded because they did not meet these criteria). In response to Baron and Goodwin's (2020) critique, we reassessed the properties of the dilemmas in the final battery using the data of the eight studies reported by Gawronski et al. (2017). Our reanalysis revealed that the manipulation of consequences was well captured by every single item across studies (see Table 2). However, the manipulation of moral norms revealed anomalies for one of the six basic dilemmas (see Table 2). Specifically, we found that the Abduction Dilemma showed patterns of action vs. inaction responses that were unaffected by the manipulation of moral norms in six of the eight studies. Moreover, in the two studies that revealed a significant effect of moral norms, the pattern of responses was opposite to the intended manipulation. Based on these findings, we deem the Abduction Dilemma as not suitable for our modeling approach. The results obtained for our other five basic dilemmas generally supported their validity.

Given that the CNI model showed adequate fit across studies (see Table 1) and the *N* parameter was significantly different from zero in every single case except for a subsample of participants high in psychopathy (see above), we deem it unlikely that the poor validity of the Abduction Dilemma affected any of our findings. Nevertheless, to gain greater confidence in the reliability of our findings, we reanalyzed the data of all eight studies reported in Gawronski et al. (2017) after excluding responses to the four variants of the Abduction Dilemma.[2] The original results reported by Gawronski et al. (2017) are summarized in Table 3, along with the results of our reanalysis. Overall, the results of the reanalysis were highly consistent with our original findings. Of the 24 comparisons, 23 revealed equivalent results in terms of whether the comparison reached statistical significance with an alpha criterion of $p < .05$. The only difference was obtained for a significant effect in the original analysis that turned marginal in the reanalysis. This comparison involves a greater sensitivity to consequences among women compared to men in Study 1b, an effect that we did not interpret in the original article, because it did not emerge in Study 1a. Together, these results suggest that the inclusion of the Abduction Dilemma did not affect the findings reported in Gawronski et al. (2017). Nevertheless, based on the results of our item analysis, we recommend that researchers using the CNI model exclude the Abduction Dilemma from the pool of items. To compensate for the smaller number of items (and the resulting loss of statistical power), researchers may

[2]The analysis files for the reported reanalyses are available at https://osf.io/u59zs/.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

TABLE 2: Proportion of "action" responses as a function of consequences (benefits of action greater vs. smaller than costs), moral norms (proscriptive vs. prescriptive), and dilemma in the studies by Gawronski, Armstrong, Conway, Friesdorf, and Hütter (2017).

	Proscriptive Norm		Prescriptive Norm		Consequences Effect		Moral Norms Effect	
	Benefits Greater than Cost	Benefits Smaller than Cost	Benefits Greater than Cost	Benefits Smaller than Cost	*F*	*p*	*F*	*p*
Dilemma 1: Abduction								
Study 1a	66.7%	43.8%	60.2%	27.9%	73.99	<.001	5.77	.017
Study 1b	59.4%	39.6%	68.5%	36.5%	63.43	<.001	0.45	.504
Study 2a	65.5%	44.8%	66.0%	24.2%	82.98	<.001	5.50	.020
Study 2b	65.5%	42.8%	64.4%	27.3%	74.94	<.001	3.32	.070
Study 3a	67.7%	31.2%	69.9%	31.2%	126.08	<.001	0.05	.818
Study 3b	64.6%	36.0%	67.7%	29.1%	91.33	<.001	0.16	.686
Study 4a	62.0%	39.7%	66.8%	23.9%	90.79	<.001	1.21	.274
Study 4b	60.6%	39.9%	62.1%	24.2%	71.20	<.001	2.47	.118
Dilemma 2: Transplant								
Study 1a	26.4%	19.9%	77.6%	67.2%	13.65	<.001	115.09	<.001
Study 1b	18.3%	12.7%	82.2%	74.4%	8.50	.004	304.55	<.001
Study 2a	18.6%	14.4%	81.4%	70.6%	12.94	<.001	219.41	<.001
Study 2b	18.0%	16.5%	74.2%	70.1%	1.99	.160	161.69	<.001
Study 3a	24.2%	18.3%	75.8%	71.5%	7.99	.005	109.13	<.001
Study 3b	19.7%	18.1%	81.9%	77.7%	3.72	.055	175.01	<.001
Study 4a	24.5%	20.7%	73.4%	72.8%	1.28	.259	97.20	<.001
Study 4b	26.3%	27.3%	67.7%	62.6%	0.80	.372	55.35	<.001
Dilemma 3: Torture								
Study 1a	65.7%	16.4%	68.2%	44.3%	128.37	<.001	13.01	<.001
Study 1b	67.0%	14.7%	76.6%	51.8%	127.19	<.001	34.09	<.001
Study 2a	71.1%	17.5%	61.3%	49.5%	96.74	<.001	7.09	.008
Study 2b	68.6%	17.5%	68.0%	48.5%	96.12	<.001	15.12	<.001
Study 3a	62.4%	8.1%	74.7%	48.4%	137.27	<.001	38.46	<.001
Study 3b	73.5%	13.8%	67.2%	40.7%	202.64	<.001	5.20	.024
Study 4a	72.8%	21.7%	59.8%	23.9%	133.39	<.001	1.63	.203
Study 4b	67.2%	22.2%	58.1%	35.4%	104.28	<.001	0.20	.653

Continued.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

	Proscriptive Norm		Prescriptive Norm		Consequences Effect		Moral Norms Effect	
	Benefits Greater than Cost	Benefits Smaller than Cost	Benefits Greater than Cost	Benefits Smaller than Cost	*F*	*p*	*F*	*p*
Dilemma 4: Assisted Suicide								
Study 1a	64.7%	33.8%	73.6&	49.3%	91.69	<.001	8.56	.004
Study 1b	62.9%	31.5%	75.6%	50.3%	78.83	<.001	16.31	<.001
Study 2a	64.4%	34.0%	72.7%	53.6%	63.45	<.001	11.29	.001
Study 2b	61.3%	27.3%	75.8%	51.0%	93.01	<.001	22.91	<.001
Study 3a	59.1%	25.3%	68.8%	44.6%	73.76	<.001	12.07	.001
Study 3b	60.8%	21.2%	67.7%	41.3%	114.01	<.001	12.61	<.001
Study 4a	54.3%	23.4%	66.3%	47.3%	58.49	<.001	16.27	<.001
Study 4b	59.1%	26.8%	70.2%	42.9%	81.65	<.001	10.77	.001
Dilemma 5: Immune Deficiency								
Study 1a	28.9%	24.4%	52.2%	40.3%	13.28	<.001	18.27	<.001
Study 1b	29.9%	27.9%	54.8%	43.1%	8.50	.004	21.17	<.001
Study 2a	28.4%	24.7%	54.1%	46.4%	6.73	.010	29.00	<.001
Study 2b	30.4%	30.4%	59.3%	45.9%	6.97	.009	29.93	<.001
Study 3a	19.5%	18.9%	61.1%	53.0%	4.20	.042	69.61	<.001
Study 3b	30.7%	32.3%	56.1%	47.1%	2.47	.118	20.13	<.001
Study 4a	28.3%	27.2%	53.8%	43.5%	4.95	.027	22.10	<.001
Study 4b	27.3%	26.3%	57.1%	45.5%	6.77	.010	36.08	<.001
Dilemma 6: Vaccine								
Study 1a	53.7%	36.8%	75.1%	65.7%	28.23	<.001	26.55	<.001
Study 1b	50.3%	35.5%	80.7%	72.1%	16.54	<.001	59.87	<.001
Study 2a	60.3%	37.6%	75.8%	69.6%	27.90	<.001	23.11	<.001
Study 2b	52.5%	37.6%	77.8%	71.6%	15.16	<.001	45.24	<.001
Study 3a	48.4%	24.2%	79.0%	66.1%	41.03	<.001	63.16	<.001
Study 3b	59.3%	34.4%	76.7%	60.8%	42.85	<.001	21.11	<.001
Study 4a	64.1%	34.8%	68.5%	47.8%	59.28	<.001	2.78	.097
Study 4b	57.6%	31.8%	68.2%	50.5%	48.85	<.001	9.22	.003

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

TABLE 3: Estimated parameter scores for sensitivity to consequences (*C*), sensitivity to moral norms (*N*), and general preference for inaction versus action (*I*) as a function of gender (Studies 1a and 1b), cognitive load (Studies 2a and 2b), question framing (Studies 3a and 3b), and psychopathy (Studies 4a and 4b). The table presents the original results reported by Gawronski, Armstrong, Conway, Friesdorf and Hütter (2017), and results with a reduced dilemma set that does not include the abduction dilemma.

Full Dilemma Set							Without Abduction Dilemma						
Study	Score	95% CI	Score	95% CI	$G^2(1)$	p	Study	Score	95% CI	Score	95% CI	$G^2(1)$	p
1a	men		women		difference		1a	men		women		difference	
C	.19	[.15, .23]	.22	[.18, .26]	1.34	.247	*C*	.18	[.14, .22]	.20	[.15, .24]	0.28	.595
N	.15	[.10, .19]	.33	[.28, .38]	26.00	<.001	*N*	.23	[.18, .28]	.37	[.32, .43]	13.63	<.001
I	.48	[.45, .51]	.56	[.52, .60]	12.34	<.001	*I*	.48	[.45, .51]	.57	[.53, .61]	11.47	<.001
1b	men		women		difference		1b	men		women		difference	
C	.16	[.12, .20]	.23	[.19, .27]	6.43	.011	*C*	.16	[.11, .20]	.21	[.17, .25]	3.22	.073
N	.26	[.21, .30]	.40	[.35, .45]	18.95	<.001	*N*	.32	[.27, .37]	.45	[.39, .50]	15.91	.001
I	.46	[.42, .49]	.53	[.49, .57]	9.12	.003	*I*	.46	[.42, .50]	.53	[.49, .48]	6.10	.013
2a	low load		high load		difference		2a	low load		high load		difference	
C	.21	[.17, .25]	.18	[.14, .22]	1.35	.245	*C*	.19	[.14, .23]	.15	[.11, .20]	1.31	.253
N	.25	[.20, .30]	.25	[.20, .30]	0.01	.927	*N*	.32	[.27, .37]	.31	[.26, .37]	0.03	.852
I	.47	[.44, .50]	.52	[.49, .56]	5.19	.023	*I*	.47	[.44, .51]	.53	[.49, .56]	3.98	.046
2b	low load		high load		difference		2b	low load		high load		difference	
C	.21	[.17, .25]	.17	[.13, .21]	2.08	.149	*C*	.19	[.15, .23]	.15	[.10, .19]	2.08	.149
N	.28	[.23, .32]	.27	[.22, .32]	0.05	.826	*N*	.35	[.30, .40]	.33	[.27, .38]	0.41	.521
I	.46	[.43, .49]	.55	[.51, .58]	13.77	<.001	*I*	.45	[.41, .49]	.56	[.52, .59]	14.18	<.001
3a	moral judgment		moral action		difference		3a	moral judgment		moral action		difference	
C	.20	[.16, .24]	.25	[.21, .28]	2.44	.118	*C*	.17	[.16, .21]	.22	[.18, .26]	3.37	.068
N	.39	[.34, .44]	.33	[.28, .38]	3.14	.069	*N*	.44	[.39, .49]	.39	[.33, .44]	2.10	.147
I	.48	[.44, .52]	.61	[.57, .65]	23.25	<.001	*I*	.49	[.48, .54]	.62	[.58, .67]	17.45	<.001
3b	moral judgment		moral action		difference		3b	moral judgment		moral action		difference	
C	.23	[.19, .27]	.22	[19, .26]	0.09	.767	*C*	.21	[.16, .25]	.21	[.17, .25]	<0.01	.970
N	.32	[.26, .37]	.23	[18., 28]	6.15	.013	*N*	.37	[.31, .42]	.28	[.22, .33]	5.32	.021
I	.44	[41, .48]	.58	[.55, .61]	29.50	<.001	*I*	.44	[.40, .48]	.59	[.55, .63]	27.00	<.001
4a	low psychopathy		high psychopathy		difference		4a	low psychopathy		high psychopathy		difference	
C	.25	[.21, .29]	.20	[.16, .24]	2.77	.096	*C*	.23	[.19, .27]	.18	[.13, .22]	2.71	.100
N	.25	[.20, .30]	.12	[.07, .17]	12.35	<.001	*N*	.33	[.27, .38]	.14	[.09, .19]	22.74	<.001
I	.57	[.54, .61]	.53	[.50, .56]	3.15	.076	*I*	.58	[.54, .63]	.54	[.51, .57]	2.71	.100
4b	low psychopathy		high psychopathy		difference		4b	low psychopathy		high psychopathy		difference	
C	.27	[.23, .30]	.13	[.09, .17]	23.11	<.001	*C*	.24	[.20, .28]	.12	[.08, .17]	15.27	<.001
N	.36	[.32, .41]	.00	[-.05, .05]	111.80	<.001	*N*	.45	[.40, .50]	.00	[-.05, .05]	148.20	<.001
I	.59	[.56, .63]	.53	[.50, .55]	8.90	.003	*I*	.62	[.57, .66]	.52	[.49, .55]	13.60	<.001

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

include the dilemmas of a newly created battery by Körner, Deutsch and Gawronski (2020).[3]

6 "Perverse" Responses

In the preceding section, we responded to Baron and Goodwin's (2020) concern that "perverse" responses on a subset of CNI model dilemmas suggest problems with the validity of these dilemmas. As we explained, the validity of a given dilemma for our modeling approach depends, not on the presence versus absence of "perverse" responses, but on the sensitivity of the four variants of a given dilemma in capturing the manipulations of consequences and moral norms. However, the mere occurrence of "perverse" responses led Baron and Goodwin to raise an additional question: what leads to "perverse" responses to our CNI model dilemmas? In their critique, Baron and Goodwin suggested two potential answers: (1) participants disagree with our assumptions underlying the manipulations of consequences and moral norms and (2) participants pay insufficient attention.

We already responded to Baron and Goodwin's (2020) first answer in the section entitled "Moral Norms." Regarding their second answer, we would like to reiterate that the mere occurrence of "perverse" responses is irrelevant for our modeling approach. What is relevant instead is whether responses differ as a function of our manipulations of consequences and moral norms. From this perspective, Baron and Goodwin's question about the causes of "perverse" responses must be rephrased as: what makes participants insensitive to the manipulations of consequences and moral norms?

We fully agree that lack of attention can be an important factor in this regard. After all, not reading the stimulus materials will reduce effects of any experimental manipulation that is embedded in the verbal materials participants are asked to read, and this caveat applies to any study using text-based manipulations regardless of whether they involve the CNI model. For this reason, we generally include instructional attention checks to identify participants who may not have read the dilemmas (see Oppenheimer, Meyvis & Davidenko, 2009). Such attention checks are particularly important for analyses regarding associations between response times and dilemma judgments (such as the ones presented by Baron and Goodwin), because inattentiveness can produce artifacts resulting from lower response times among participants who do not carefully read the dilemmas. Indeed, if inattentiveness can be ruled out via attention checks, associations between response times and the CNI model parameters may indicate meaningful relations between cognitive elaboration and moral judgments. However, without proper attention checks, any such associations may be driven by lack of attention, as suggested by Baron and Goodwin (2020). We generally agree with this concern.

Aside from inattentiveness, there are multiple other factors that can reduce participants' sensitivity to the manipulations of consequences and moral norms. However, different from Baron and Goodwin's claim that any such effects demonstrate problems with our modeling approach, we argue that they reflect meaningful variations in moral judgments. For example, as we noted above, the only case in which the *N* parameter did not significantly differ from zero involved a subsample of participants who were preselected based on having psychopathy scores that placed them in the highest quartile of participants in a broader sample (Gawronski et al., 2017, Study 4b). If one were to follow Baron and Goodwin's arguments, this result would be meaningless, because it suggests that these participants disagree with our assumptions underlying the manipulation of moral norms. However, based on the known characteristics of individuals high in psychopathy (Hare & Neumann, 2008), a more plausible interpretation is that individuals high in psychopathy are less sensitive to moral norms than individual low in psychopathy. This conclusion is consistent with the difference between participants high versus low in psychopathy obtained by Gawronski et al. (2017, Study 4b) and successful replications of this effect in several follow-up studies (Körner et al., 2020; Luke & Gawronski, in press). In our view, these results do not call the validity of the CNI model into question. Instead, they demonstrate the validity of the model and its ability to provide more nuanced insights into central questions at the intersection of moral and clinical psychology. We will return to these findings in the section entitled "Confounds in the Traditional Approach."

Another argument put forward by Baron and Goodwin is that the proportion of "perverse" responses is reduced if participants' own judgments are used to identify the relevant moral norm in a given dilemma. As we explained above, we deem this approach problematic, because it conflates two distinct levels of analysis (see De Houwer, 2011; Gawronski & Bodenhausen, 2015). In fact, classification of a behavioral effect by means of underlying mental constructs would lead to explanatory circularity, because the to-be-explained effect becomes conceptually equivalent with the construct that is supposed to explain the effect (see De Houwer, Gawronski & Barnes-Holmes, 2013; Gawronski & Bodenhausen, 2015). That is, if norm-congruent behavior would qualify as an effect of norms only if it is mediated by conscious thoughts about norms, conscious thoughts about norms would not explain norm-congruent behavior, because the two would be conceptually identical. It makes little sense to *explain* norm-congruent behavior by conscious thoughts about norms if norm-congruent behavior is *defined* by conscious thoughts about norms. Such a conceptualization would also preclude the possibility that norm-congruent behavior might be mediated by other mental processes that do not involve con-

[3] A reanalysis of the data by Körner et al. (2020) revealed that every item of their new dilemma battery meets the specified validity criteria for capturing effects of consequences and moral norms.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

scious thoughts about norms, such as the affective mechanisms proposed by Greene's (2008) dual-process theory (see also Haidt, 2001).

For these reasons, we reject Baron and Goodwin's idea of using participants' self-classifications as a criterion for moral norms. Instead, we would argue that the mechanisms underlying norm-congruent judgments is an empirical question. It might be conscious thoughts about moral norms, affective responses to the idea of causing harm, or something entirely different. In fact, it is possible that multiple distinct processes can produce the same pattern of norm-congruent judgments. Similar considerations apply to effects of consequences, which may be driven by conscious reasoning about consequences or an entirely different set of processes. In hindsight, we acknowledge that the description of the *N* parameter as an indicator of sensitivity to moral norms is not ideal, because it is ambiguous about the difference between behavioral effects and underlying mental processes. A less ambiguous description might refer to intrinsic aspects of the focal action (e.g., actions causing harm vs. actions preventing harm; see Cushman, 2013) rather than the congruency of these actions with proscriptive and prescriptive norms. Such a description would also address the abovementioned concern that congruence with moral norms is not a unique feature of the *N* parameter (see discussion under "Deontological Responding"). Yet, regardless of the chosen label, the significance of the *N* parameter in understanding moral dilemma judgments is supported by the finding that the *N* parameter explains a considerable amount of variance in traditional dilemma scores and accounts for a large number of experimental effects and associations with individual-difference variables. We will return to this issue in the section entitled "Confounds in the Traditional Approach."

7 Order Effects and Presumed Inconsistencies

Multinomial modeling is not just an alternative way of analyzing data. Different from standard data analytical approaches (e.g., General Linear Model), multinomial modeling integrates data analysis with theoretical assumptions about observed response patterns (Hütter & Klauer, 2016). In the CNI model, these assumptions are reflected in the hierarchical relation of the three parameters in the processing tree (see Figure 1), which can be criticized as arbitrary, because they cannot be tested empirically. As we explained in the original article presenting the CNI model (Gawronski et al., 2017, Footnote 7), changes in the order of parameters in the processing tree do not affect the fit of the model, in that all six combinatorially possible models show the same goodness-of-fit for a given data set. It is therefore not possible to empirically distinguish between these models, rendering assumptions about the hierarchical structure of the three parameters arbitrary.

Baron and Goodwin express two concerns about this aspect of the CNI model. First, they claim that the order of *C* and *N* in the processing tree conflicts with the corrective dual-process theory (Greene, 2008), which suggests that deontological judgments are the product of automatic processes, whereas utilitarian judgments are the product of controlled processes that correct the impact of automatically generated deontological judgments. Second, they claim that models with different orders of *C* and *N* in the processing tree lead to inconsistent relations with external variables. Both claims are incorrect.

Regarding the consistency of the parameter hierarchy with the corrective dual-process theory (Greene, 2008), it is important to note that the hierarchical structure reflects conditional relations of the parameters in determining behavioral outcomes, not their temporal order in which their underlying processes occur. In technical terms, the hierarchical structure of parameters specifies conditional probabilities of input-output relations, and time is not a meaningful variable in the mathematics of conditional probabilities. In the CNI model, we chose the current order of *C* and *N* to be consistent with Conway and Gawronski's (2013) earlier work using process dissociation, and the order of the corresponding parameters in their process dissociation model was chosen precisely because it is the one suggested by the corrective dual-process theory. According to the theory, the processes underlying deontological judgments influence outcomes only if they are not overridden by the corrective processes underlying utilitarian judgments. In other words, the corrective processes underlying utilitarian judgments will drive responses whenever these processes become active, and the processes underlying deontological judgments will drive responses only if the corrective processes underlying utilitarian judgments fail to drive judgments. In Conway and Gawronski's (2013) process dissociation model, these assumptions are reflected in the dominant status of the *U* parameter (claimed to reflect utilitarian inclinations) compared to the subordinate status of the *D* parameter (claimed to reflect deontological inclinations) in the hierarchy of parameters. Within the CNI model, these assumptions are retained in the dominant status of the *C* parameter compared to the subordinate status of the *N* parameter. As we explained in the original article presenting the CNI model (Gawronski et al., 2017, Footnote 7), the position of the *I* parameter was chosen to permit an estimation of general action preferences along a bipolar continuum of inaction versus action instead of a unipolar dimensions reflecting relative differences in the general preference for inaction. This aspect requires that the *I* parameter is included at the lowest location in the hierarchy instead of being at one of the two superordinate locations.

Although the position of the *I* parameter can be justified on methodological grounds, we agree with Baron and Goodwin

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

(2020) that the relative position of *C* and *N* in the processing tree remains arbitrary, raising the question of whether reversing their positions qualifies any of the reported results. Mathematically, reversing the positions of *C* and *N* changes the model equations, in that the score estimated for *N* is conditional upon $1 - C$ in the original model, whereas the score estimated for *C* is conditional upon $1 - N$ in the reversed model (see Appendix B). Although these changes do not affect the goodness-of-fit of the model (see above), they can influence relations with external variables. Thus, it is perfectly justified to ask the question of whether the findings obtained with the CNI model depend on the chosen position of *C* and *N* in the processing tree.

Baron and Goodwin (2020) claim that reordering *C* and *N* in the processing tree leads to inconsistent results, with some external variables (e.g., gender) showing relations to the parameters of the reversed model that are opposite to the ones obtained for the original model. However, a closer inspection of the results indicates that this claim is incorrect. Table 4 provides a summary of the results obtained with the original model (C-N-I order), along with a summary of the results obtained with a model in which the positions of *C* and *N* are reversed (N-C-I order). For 21 of the 24 comparisons, the two models produce the same outcomes in terms of whether the comparison reached statistical significance with an alpha criterion of $p < .05$. In two cases, a marginal difference obtained with the original model reaches statistical significance with the reversed model. These cases involve the effect of question framing on the *N* parameter in Study 3a and the association of psychopathy with the *C* parameter in Study 4a. In one case, a non-significant difference in the original model is statistically significant in the reversed model. This case involves the association between gender and the *C* parameter in Study 1a. Counter to Baron and Goodwin's (2020) claim, all of the obtained relations between the three model parameters and external variables are directionally consistent across the two model variants.[4]

How is it possible that Baron and Goodwin (2020) reached an entirely different conclusion in their own reanalysis of our data? The simple reason is that they did not use the CNI model in their reanalysis. Instead, they used two versions of Conway and Gawronski's (2013) process dissociation model, one in which *U* dominates over *D* (as in Conway and Gawronski' model) and one in which *D* dominates over *U* (opposite to the order in Conway and Gawronski's model). Baron and Goodwin then used the two versions to separately analyze responses to dilemmas with either proscriptive or prescriptive norms. Using this approach to reanalyze the data from Studies 1a and 1b of Gawronski et al. (2017), Baron and Goodwin identified inconsistencies in the relation between gender and the *U* parameter across the four cases, with one case showing gender effects that are directionally opposite to the ones in the other three cases.

We deem these findings uninformative for the validity of the CNI model, because (1) the analyses did not use the CNI model and (2) a proper analysis with the CNI model yields consistent results across model specifications (see Table 4). Indeed, there is a simple technical explanation why Baron and Goodwin's (2020) reanalysis using the PD model to compare results across dilemmas with proscriptive and prescriptive norms yields inconsistent results. In the PD model, general preference for inaction versus action is confounded with the two constructs of interest (see Gawronski et al., 2016; Hütter & Klauer, 2016), and these confounds are directionally opposite in cases involving proscriptive and prescriptive norms. For dilemmas involving proscriptive norms, a general preference for inaction versus action increases scores on the *D* parameter and decreases scores on the *U* parameter. For dilemmas involving prescriptive norms, a general preference for inaction versus action decreases scores on the *D* parameter and increases scores on the *U* parameter. Because the relative impact of these confounds varies as a function of the hierarchy of the two parameters in the processing tree, they can lead to inconsistent results as a function of model specifications (i.e., U-D order vs. D-U order) and type of norm (proscriptive vs. prescriptive) if a given variable (e.g., gender) is associated with general preference for inaction versus action. By controlling for general action preferences in the *I* parameter, the CNI model not only resolves these problems; it also leads to consistent results irrespective of the position of *C* and the *N* in the processing tree (see Table 4). Thus, although the results of Baron and Goodwin's (2020) reanalysis echo our own concerns about conceptual problems of Conway and Gawronski's (2013) PD model (see Gawronski et al., 2016; Hütter & Klauer, 2016), they have no implications for the validity of the CNI model.

8 Confounds in the Traditional Approach

Baron and Goodwin (2020) conclude their critique stating that the "CNI model requires use of congruent items that must yield enough 'perverse' responses (those that both violate norms and produce worse consequences) so that the model provides results that differ from standard analysis" (p. 434). As we explained above, their focus on 'perverse' responses reflects a misunderstanding of the model's underlying idea. Whereas the *C* parameter quantifies the extent to

[4]To further investigate the robustness of our original findings, we have also used the reversed model to reanalyze Gawronski et al.'s (2017) data without the Abduction Dilemma. The analyses yielded similar results, in that the two models produced the same outcomes for 22 of the 24 comparisons. In the two deviating cases, a marginal difference obtained with the original model reaches statistical significance with the reversed model. These cases involve the effect of question framing on the *N* parameter in Study 3a and the association between psychopathy and the *C* parameter in Study 4a. The results of this reanalysis are summarized in Appendix C. The analysis files for the reported reanalyses are available at https://osf.io/u59zs/.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

TABLE 4: Estimated parameter scores for sensitivity to consequences (*C*), sensitivity to moral norms (*N*), and general preference for inaction versus action (*I*) as a function of gender (Studies 1a and 1b), cognitive load (Studies 2a and 2b), question framing (Studies 3a and 3b), and psychopathy (Studies 4a and 4b). The table presents the results of Gawronski, Armstrong, Conway, Friesdorf and Hütter's (2017) studies using the original CNI model (C-N-I order) and a modified version in which the order of N and C are reversed in the hierarchical structure of the model (N-C-I order).

C-N-I Order							N-C-I Order						
Study	Score	95% CI	Score	95% CI	$G^2(1)$	p	Study	Score	95% CI	Score	95% CI	$G^2(1)$	p
1a	men		women		difference		1a	men		women		difference	
C	.19	[.15, .23]	.22	[.18, .26]	1.34	.247	*C*	.21	[.17, .27]	.29	[.24, .35]	5.51	.019
N	.15	[.10, .19]	.33	[.28, .38]	26.00	<.001	*N*	.12	[.08, .16]	.25	[.22, .29]	23.74	<.001
I	.48	[.45, .51]	.56	[.52, .60]	12.34	<.001	*I*	.48	[.45, .51]	.56	[.52, .60]	12.34	<.001
1b	men		women		difference		1b	men		women		difference	
C	.16	[.12, .20]	.23	[.19, .27]	6.43	.011	*C*	.21	[.16, .26]	.33	[.28, 39]	11.77	<.001
N	.26	[.21, .30]	.40	[.35, .45]	18.95	<.001	*N*	.22	[.18, .26]	.31	[.27, .35]	11.61	<.001
I	.46	[.42, .49]	.53	[.49, .57]	9.12	.003	*I*	.46	[.42, .49]	.53	[.49, .57]	9.12	.003
2a	low load		high load		difference		2a	low load		high load		difference	
C	.21	[.17, .25]	.18	[.14, .22]	1.35	.245	*C*	.26	[.21, .31]	.22	[.17, .27]	1.12	.289
N	.25	[.20, .30]	.25	[.20, .30]	0.01	.927	*N*	.20	[.16, .24]	.21	[.17, .25]	0.15	.700
I	.47	[.44, .50]	.52	[.49, .56]	5.19	.023	*I*	.47	[.44, .50]	.52	[.49, .56]	5.19	.023
2b	low load		high load		difference		2b	low load		high load		difference	
C	.21	[.17, .25]	.17	[.13, .21]	2.08	.149	*C*	.27	[.22, .32]	.22	[.16, .27]	1.93	.164
N	.28	[.23, .32]	.27	[.22, .32]	0.05	.826	*N*	.22	[.18, .26]	.22	[.18, .26]	0.03	.864
I	.46	[.43, .49]	.55	[.51, .58]	13.77	<.001	*I*	.46	[.43, .49]	.55	[.51, 58]	13.77	<.001
3a	moral judgment		moral action		difference		3a	moral judgment		moral action		difference	
C	.20	[.16, .24]	.25	[.21, .28]	2.44	.118	*C*	.30	[.24, .35]	.33	[.28, .38]	0.66	.418
N	.39	[.34, .44]	.33	[.28, .38]	3.14	.069	*N*	.31	[.27, .35]	.25	[.21, .28]	5.78	.016
I	.48	[.44, .52]	.61	[.57, .65]	23.25	<.001	*I*	.48	[.44, .52]	.61	[.57, .65]	23.25	<.001
3b	moral judgment		moral action		difference		3b	moral judgment		moral action		difference	
C	.23	[.19, .27]	.22	[19, .26]	0.09	.767	*C*	.31	[.25, .36]	.27	[.22, .32]	0.96	.326
N	.32	[.26, .37]	.23	[18., 28]	6.15	.013	*N*	.24	[.20, .28]	.18	[.14, .21]	5.85	.016
I	.44	[41, .48]	.58	[.55, .61]	29.50	<.001	*I*	.44	[.41, .48]	.58	[.55, .61]	29.50	<.001
4a	low psychopathy		high psychopathy		difference		4a	low psychopathy		high psychopathy		difference	
C	.25	[.21, .29]	.20	[.16, .24]	2.77	.096	*C*	.30	[.26, .35]	.22	[.17, .27]	6.08	.013
N	.25	[.20, .30]	.12	[.07, .17]	12.35	<.001	*N*	.19	[.15, .23]	.10	[.06, .14]	10.35	.001
I	.57	[.54, .61]	.53	[.50, .56]	3.15	.076	*I*	.57	[.54, .61]	.53	[.50, .56]	3.15	.076
4b	low psychopathy		high psychopathy		difference		4b	low psychopathy		high psychopathy		difference	
C	.27	[.23, .30]	.13	[.09, .17]	23.11	<.001	*C*	.36	[.31, .41]	.13	[.09, .17]	49.43	<.001
N	.36	[.32, .41]	.00	[-.05, .05]	111.80	<.001	*N*	.27	[.23, 30]	.00	[-.04, .04]	95.34	<.001
I	.59	[.56, .63]	.53	[.50, .55]	8.90	.003	*I*	.59	[.56, .63]	.53	[.50, .55]	8.90	.003

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

which responses are influenced by consequences and the *N* parameter quantifies the extent to which responses are influenced by moral norms, the *I* parameter quantifies the extent to which responses reflect a general preference for inaction versus action. In technical terms, the *C* and the *N* parameter quantify the impact of two experimental manipulations; the *I* parameter is conceptually similar to an intercept in a General Linear Model, in that it captures general response tendencies independent of the two experimental manipulations. Thus, what matters within the CNI model are differences (or the lack of differences) in responses across dilemmas involving different consequences and moral norms, not absolute responses to a particular kind of dilemma. In fact, counter Baron and Goodwin's claim that the CNI model requires "perverse" responses, substantial proportions of "perverse" responses would be detrimental to the model, because they would lead to (1) poor model fit in the description of data due to violations of model assumptions and (2) estimates for the *C* and the *N* parameters close to zero due to reduced effects of their corresponding manipulations.

Counter to Baron and Goodwin's (2020) preference for the traditional approach of pitting moral norms against consequences for the greater good, we argue that the three factors captured by the CNI model are confounded in the traditional approach. These confounds render any finding with the traditional approach ambiguous, in that differences in the preference for "utilitarian" over "deontological" judgments may be driven by differences in the sensitivity to consequences, differences in the sensitivity to moral norms, or differences in the general preference for inaction versus action (or any combination of the three).

The conflation of the three factors in the traditional approach can be illustrated by means of multiple regression analyses, in which preference for "utilitarian" over "deontological" judgments on traditional dilemmas pitting a proscriptive norm against consequences for the greater good is predicted by the three CNI parameters. Table 5 presents the results of such multiple regression analyses for a series of four studies by Körner et al. (2020).[5] Because responses on traditional dilemmas are used in the CNI model equations to estimate numerical values for the three parameters (see Appendix A), we ensured mathematical independence of predictors and outcomes by using CNI model parameters for dilemmas with odd item numbers to predict traditional dilemma scores for dilemmas with even item numbers. Conversely, we used CNI model parameters for dilemmas with even item numbers to predict traditional dilemma scores for dilemmas with odd item numbers. For the four studies reported by Körner et al. (2020), this approach led a total of eight multiple regressions. In Studies 1a and 1b, participants were asked to judge whether the described action is acceptable. In Studies 2a and 2b, participants were asked whether they would perform the described action.[6]

Table 5 shows the results of the multiple regression analyses. For the *N* parameter, greater sensitivity to moral norms showed a significant negative association with preference for "utilitarian" over "deontological" judgments in all eight cases. For the *C* parameter, greater sensitivity to consequences showed a significant positive association with preference for "utilitarian" over "deontological" judgments in seven of the eight regressions. For the *I* parameter, the results were somewhat less reliable, in that general preference for inaction versus action showed a significant negative association with preference for "utilitarian" over "deontological" judgments in four of the eight regressions. A potential reason for the mixed results for the *I* parameter is that estimates for parameters with a lower position in the hierarchy of multinomial processing trees tend to be less reliable compared to estimates for parameters with a higher position in the hierarchy.[7] Nevertheless, the regression results indicate that the three CNI parameters capture unique variance in traditional dilemmas scores and the direction of their associations is consistent with the proposed confounds.

By disentangling the three determinants of moral dilemma judgments, the CNI model provides much more nuanced insights that cannot be gained with the traditional approach. For example, although reduced preference for "utilitarian" over "deontological" judgments under cognitive load has been interpreted as reflecting the resource-dependence of utilitarian reasoning about costs and benefits (Greene, Morelli, Lowenberg, Nystrom & Cohen, 2008), results obtained with the CNI model suggest that cognitive load influences moral dilemma judgments by increasing general preference for inaction versus action (Gawronski et al., 2017, Studies 2a and 2b). Moreover, Gawronski and Brannon (2020) found that greater preference for "utilitarian" over "deontological" judgments resulting from recalling personal memories of high (vs. low) power is driven by a reduced

[5]Because multinomial modeling requires a sufficient number of observations from each participant to provide reliable parameter estimates at the individual level, such regression analyses were not feasible with the original set of 24 dilemmas for research using the CNI model (Gawronski et al., 2017). However, estimation of parameters at the individual level is possible with an extended set of 48 dilemmas presented by Körner et al. (2020), which was used for the analyses reported in Table 5. The analysis files for the reported reanalyses are available at https://osf.io/u59zs/.

[6]In Footnote 9 of their critique, Baron and Goodwin (2020) express concerns that asking participants about the acceptability of the described action may induce a bias against utilitarian responding. In response to this concern, it is worth noting that asking participants whether they would perform the described action leads to higher mean-level scores on the *I* parameter and lower mean-level scores on the *N* parameter, with scores on the *C* parameter being unaffected by the type of question framing (see Gawronski et al., 2017; Körner et al., 2020). Nevertheless, relations to external variables have been found to be invariant for the two kinds of question framings (Körner et al., 2020), rendering the obtained mean-level differences inconsequential for research investigating effects of experimental manipulations and associations with individual-difference variables.

[7]Consistent with this interpretation, Spearman-Brown coefficients based on test halves with odd and even trials were .65 for the *C* parameter, .67 for the *N* parameter, and .35 for the *I* parameter in the combined data of the four studies.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

TABLE 5: Results of multiple regression analyses predicting traditional dilemma scores (i.e., preference for "utilitarian" over "deontological" judgments on dilemmas that pit a proscriptive norm against consequences for the greater good) by sensitivity to consequences (*C*), sensitivity to moral norms (*N*), and general preference for inaction versus action (*I*). For CNI parameters marked *odd*, the CNI parameters were calculated based on dilemmas with odd trial numbers and traditional dilemma scores were calculated based on dilemmas with even trial numbers. For CNI parameters marked *even*, the CNI parameters were calculated based on dilemmas with even trial numbers and traditional dilemma scores were calculated based on dilemmas with odd trial numbers. Reanalysis of data by Körner, Deutsch, and Gawronski (2020).

	Study 1a N = 161			Study 1b N = 177			Study 2a N = 196			Study 2b N = 189		
	β	t	p	β	t	p	β	t	p	β	t	p
C_{odd}	.192	2.53	.012	.137	2.03	.044	.090	1.32	.187	.177	2.53	.012
N_{odd}	-.342	-4.68	<.001	-.442	-6.68	<.001	-.387	-5.79	<.001	-.340	-5.13	<.001
I_{odd}	-.274	-3.63	<.001	-.268	-4.03	<.001	-.098	-1.42	.158	-.291	-4.16	<.001
Adj. R^2	.181			.254			.160			.179		
C_{even}	.320	4.46	<.001	.287	4.35	<.001	.393	6.12	<.001	.309	4.67	<.001
N_{even}	-.383	-5.30	<.001	-.356	-5.36	<.001	-.386	-5.97	<.001	-.318	-4.70	<.001
I_{even}	-.102	-1.44	.151	-.302	-4.62	<.001	.043	0.68	.499	-.096	-1.42	.158
Adj. R^2	.208			.266			.231			.186		

sensitivity to moral norms rather than increased sensitivity to consequences or reduced action aversion. Findings by Gawronski et al. (2018) further indicate that happiness increases the preference for "utilitarian" over "deontological" judgments (see Valdesolo & DeSteno, 2006) by reducing sensitivity to moral norms, suggesting that happiness influences moral judgments by damping negative affective responses to the idea of violating moral norms (see Nichols & Mallon, 2006) rather than negative affective responses to the idea of causing harm (see Greene, 2008). Finally, Gawronski et al. (2017, Studies 3a and 3b) found that personal involvement decreased sensitivity to moral norms and increased general preference for inaction versus action (see also Körner et al., 2020), indicating that the same factor can have opposite effects on the two patterns of deontological responding.

In our view, the most interesting insights are provided by studies that have used the CNI model to investigate associations between psychopathy and moral dilemma judgments (e.g., Gawronski et al., 2017, Studies 4a and 4b; Körner et al., 2020; Luke & Gawronski, in press). Numerous previous studies using the traditional dilemma approach have found a positive association between psychopathy and preference for "utilitarian" over "deontological" judgments (for a meta-analysis, see Marshall, Watts & Lilienfeld, 2018). Some researchers have interpreted this finding as evidence for a major flaw of moral dilemma research, because it is obvious that individuals high in psychopathy do not care about the greater good in a utilitarian sense (e.g., Bartels & Pizarro, 2011; Kahane, Everett, Earp, Farias, Savulescu, 2015). Research using the CNI model resolves this paradox, showing that individuals high in psychopathy differ from those low in psychopathy by showing (1) a weaker sensitivity to consequences on the *C* parameter, (2) a weaker sensitivity to moral norms on the *N* parameter, and (3) a weaker general preference for inaction versus action on the *I* parameter. The most interesting aspect of these findings is that individuals high in psychopathy showed a weaker sensitivity to consequences in a utilitarian sense, which stands in contrast to the positive association between psychopathy and preference for "utilitarian" over "deontological" judgments in the traditional approach. This discrepancy can be explained by the finding that psychopathy shows a large negative association with sensitivity moral norms and a moderate negative association with general preference for inaction versus action, which conceal a moderate negative association with sensitivity to consequences when the three factors are confounded in the traditional dilemma approach. Further research by Luke and Gawronski (in press) suggests that some of these associations are driven by a poor understanding of societal standards about right and wrong among individuals high in psychopathy. For other associations, the results suggest that psychopaths are aware of societal standards about right and wrong, but do not care about using these standards in their personal judgments. These complex findings speak against the possibility that associations between psychopathy and the three CNI model parameters merely reflect inattentiveness among individuals high in psychopathy, as Baron and Goodwin (2020) might argue based on their critique.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

9 Conclusion

We greatly appreciate the opportunity to correct potential misconceptions about the CNI model, like the ones reflected in Baron and Goodwin's (2020) critique. Although their critique reveals a need for greater precision in the description of the three model parameters and for greater attention to properties of individual dilemmas, we deem their dismissal of the CNI model as misguided, because it is based on: (1) misunderstandings of key aspects of the model; (2) a conceptually problematic conflation of behavioral effects and explanatory mental constructs; (3) arguments that are inconsistent with empirical evidence; and (4) reanalyses that supposedly show inconsistent findings resulting from changes in model specifications, although the reported reanalyes did not actually use the CNI model and proper analyses with the CNI model yield consistent findings across model specifications. Based on the available evidence, we conclude that the CNI model is a valid, robust, and empirically sound approach to gaining deeper insights into the determinants of moral dilemma judgments, overcoming major limitations of the traditional approach.

10 References

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral judgment and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 478–515). Chichester, UK: Wiley.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121,* 154–161.

Baron, J., & Goodwin, G. P. (2020). Consequences, norms, and inaction: A critical analysis. *Judgment and Decision Making, 15*, 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,* 536–554.

Bialek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology, 85:103855*.

Brannon, S. M., Carr, S., Jin, E. S., Josephs, R. A., & Gawronski, B. (2019). Exogenous testosterone increases sensitivity to moral norms in moral dilemma judgments. *Nature Human Behavior*, *3,* 856-866.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216-235.

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.

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.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review, 17,* 273–292.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, *6*, 202–209.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology, 24,* 252–287.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5–15.

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,* 343–376.

Gawronski, B., & Beer, J. S. (2017). What makes moral dilemma judgments "utilitarian" or "deontological"? *Social Neuroscience, 12,* 626-632.

Gawronski, B., & Bodenhausen, G. V. (2015). Social-cognitive theories. In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 65– 83). New York: Guilford Press.

Gawronski, B., & Brannon, S. M. (2020). Power and moral dilemma judgments: Distinct effects of memory recall versus social roles. *Journal of Experimental Social Psychology*, *86:103908*.

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2016). Understanding responses to moral dilemmas: Deontological inclinations, utilitarian inclinations, and general action tendencies. In J. P. Forgas, L. Jussim, & P. A. M. Van Lange (Eds.), *Social psychology of morality* (pp. 91-110). New York: Psychology Press.

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,* 989-1008.

Greene, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The neuroscience of morality* (pp. 35-79). Cambridge: MIT Press.

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*, 1144–1154.

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology*, *4*, 217–246.

Hennig, M., & Hütter, M. (2020). Revisiting the divide between deontology and utilitarianism in moral dilemma judgment: A multinomial modeling approach. *Journal of Personality and Social Psychology, 118,* 22–56.

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology, 27,* 116–159.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54–69.

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.

Klauer, K. C. (2015). Mathematical modeling. In B. Gawronski, & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 371-389). New York: Guilford Press.

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

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.

Luke, D. M., & Gawronski, B. (in press). Psychopathy and moral dilemma judgments: A CNI model analysis of personal and perceived societal standards. *Social Cognition*.

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*, 40–50.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100*, 530–542.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872.

Schwarz, N. (1999). Self-reports: How the questions shape the answer. *American Psychologist, 54,* 93–105.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59,* 204–217.

Valdesolo, P. & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*, 476–477.

11 Appendix A

CNI 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. Equations adapted from Gawronski, Armstrong, Conway, Friesdorf, and Hütter (2017). Reprinted with permission from the American Psychological Association.

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

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

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

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

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

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

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

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

12 Appendix B

Model equations for the estimation of sensitivity to moral norms (N), sensitivity to consequences (C), 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. The equations characterize a model that is conceptually similar to the CNI model, the only difference being that the hierarchical positions of C and N are reversed in the processing tree (i.e., N-C-I instead of C-N-I).

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

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

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

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

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

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

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press

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

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

13 Appendix C

Estimated parameter scores for sensitivity to consequences (C), sensitivity to moral norms (N), and general preference for inaction versus action (I) as a function of gender (Studies 1a and 1b), cognitive load (Studies 2a and 2b), question framing (Studies 3a and 3b), and psychopathy (Studies 4a and 4b). The table presents the results of a reanalysis of Gawronski, Armstrong, Conway, Friesdorf, and Hütter's (2017) data with a modified version of the CNI model in which the positions of C and N are reversed in the processing tree (i.e., N-C-I instead of C-N-I) and responses to the four variants of the Abduction Dilemma are excluded.

Study	Score	95% CI	Score	95% CI	$G^2(1)$	p
1a	men		women		difference	
C	.22	[.17, .27]	.28	[.22, .34]	2.01	.156
N	.19	[.15, .23]	.30	[.26, .34]	12.85	<.001
I	.48	[.45, .51]	.57	[.53, .61]	11.47	<.001
1b	men		women		difference	
C	.21	[.16, .27]	.32	[.26, .39]	6.53	.011
N	.27	[.22, .31]	.35	[.31, .39]	8.09	.004
I	.46	[.42, .50]	.53	[.49, .58]	6.10	.013
2a	low load		high load		difference	
C	.25	[.20, .31]	.21	[.15, .27]	1.21	.272
N	.26	[.22, .30]	.27	[.22, .31]	0.03	.862
I	.47	[.44, .51]	.53	[.49, .56]	3.98	.046
2b	low load		high load		difference	
C	.26	[.21, .32]	.20	[.14, .26]	2.18	.140
N	.28	[.24, .32]	.28	[.23, .32]	0.03	.868
I	.45	[.41, .49]	.56	[.52, .59]	14.18	<.001
3a	moral judgment		moral action		difference	
C	.27	[.20, .33]	.32	[.26, .38]	1.38	.240
N	.37	[.33, .41]	.30	[.26, .34]	5.00	.025
I	.49	[.45, .54]	.62	[.58, .67]	17.45	<.001
3b	moral judgment		moral action		difference	
C	.29	[.23, .35]	.26	[.21, .32]	0.37	.543
N	.29	[.25, .33]	.22	[.18, .26]	5.51	.019
I	.44	[.40, .48]	.59	[.55, .63]	27.01	<.001
4a	low psychopathy		high psychopathy		difference	
C	.31	[.25, .36]	.20	[.15, .25]	7.41	.006
N	.25	[.21, .30]	.11	[.07, .16]	19.59	<.001
I	.58	[.54, .63]	.54	[.51, .57]	.271	.100
4b	low psychopathy		high psychopathy		difference	
C	.37	[.31, .42]	.12	[.08, .17]	41.48	<.001
N	.34	[.31, .38]	.00	[-.05, .05]	131.99	<.001
I	.62	[.57, .66]	.52	[.49, .55]	13.60	<.001

https://doi.org/10.1017/S1930297500008251 Published online by Cambridge University Press