

**CAN Algorithm: An Individual Level Approach to identify Consequences and Norms
Sensitivities and Overall Action/inaction Preferences in Moral Decision-making**

Chuanjun Liu and Jiangqun Liao

Tsinghua University

Author Note

*Correspondence concerning this article should be addressed to Jiangqun Liao, Department of Psychology, School of Social Sciences, Tsinghua University, Haidian District, Beijing, China, 100084. E-mail: liaojq@tsinghua.edu.cn. Tel & Fax: (86-10) 62787208.

This work was supported by the National Social Science Foundation of China [Grant No.18BSH114]; and Tsinghua University Initiative Scientific Research Program [Grant No. 2017THZWYY11].

ABSTRACT

Gawronski et al. (2017) developed a CNI model to measure an agent's norms sensitivity, consequences sensitivity, and generalized inaction/action preferences when making moral decisions. However, the CNI model presupposed that an agent considers *consequences*—*norms*—generalized *inaction/action* preferences sequentially, which is untenable based on recent evidence. Moreover, the CNI model generates parameters at the group level based on binary categorical data. Hence, the *C/N/I* parameters cannot be used for correlation analyses or other conventional research designs. To solve these limitations, we developed the CAN algorithm to compute *norms* and *consequences* sensitivities and overall *action/inaction* preferences algebraically in a parallel manner. We re-analyzed the raw data of Gawronski et al. (2017) to test the methodological predictions. Our results demonstrate that: (1) the *C* parameter is approximately equal between the CNI model and CAN algorithm; (2) the *N* parameter under the CNI model approximately equals $N/(1 - C)$ under the CAN algorithm; (3) the *I* parameter and *A* parameter are reversed around 0.5 – the larger the *I* parameter, the more the generalized inaction *versus* action preference and the larger the *A* parameter, the more overall action *versus* inaction preference; (4) tests of differences in parameters between groups with the CNI model and CAN algorithm led to almost the same statistical conclusion; (5) Parameters from the CAN algorithm can be used for correlational analyses and multiple comparisons, and this is an advantage over the parameters from the CNI model. The theoretical and methodological implications of our study were also discussed.

Key words: CAN algorithm; moral dilemma; moral decision-making; CNI model

INTRODUCTION

Traditional moral dilemmas pit utilitarianism against deontology. Take the well-known problem of the trolley car as an example. An uncontrollable trolley car is rushing towards five workers who do not notice this emergency. There is a sidetrack and only one worker there, not noticing this emergency either. The only way to save the five workers is to pull the switch and let the trolley car run into the sidetrack. If that occurs, the one worker on the sidetrack will die and the five workers on the main track will be saved. The principle of utilitarianism is followed if the agent chooses to pull the switch because it achieves greater benefits than costs (Bentham, 1996; Mill, 1872). The principle of deontology is followed if the agent chooses not to pull the switch because harming the innocent is not allowed according to moral norms (Kant & Gregor, 1997; Rawls, 1971).

However, interpretations of the paradigms of traditional moral dilemmas are ambiguous (Gawronski, Armstrong, Conway, Friesdorf, & Hutter, 2017; Gawronski & Beer, 2017). In the trolley-car dilemma, there could be three reasons why the agent is likely to pull the switch. The first is that the agent has weaker norm sensitivity, and is less averse to the sacrificing utilitarian proposal. The second is that the agent has a stronger consequence sensitivity, and finds the result of pulling the switch to be considerably beneficial. The third is that the agent wants to pull the switch and have a stronger generalized action (or weaker generalized inaction) preference irrespective of the norms and consequences behind it. The paradigm of the traditional dilemma cannot dissociate these three possibilities. Thus, we cannot tell whether norm sensitivity, consequence sensitivity, or generalized action/inaction preference matter in the agent's moral decision-making.

To solve this ambiguity, Gawronski et al. (2017) developed a multinomial processing tree (MPT) model to dissociate the three possible interpretations stated above. First, they expanded the conceptual manipulations of utilitarianism and deontology. They addressed the manipulation limitation of a traditional dilemma. “Utilitarian” presupposes that the observed behavior is sensitive to consequences, which requires experimental manipulations of consequences. “Deontological” presupposes that the observed behavior is sensitive to moral norms, which requires experimental manipulations of moral norms.

Hence, four types of dilemmas involving different combinations of consequences and norms must be considered (Gawronski, et al., 2017; Gawronski, & Beer, 2017). That is, dilemmas in which a: (a) *proscriptive* norm opposes the proposed behavior, and the benefits of behavior for overall wellbeing are *greater* than the costs of behavior; (b) *proscriptive* norm opposes the proposed behavior, and the benefits of behavior for overall wellbeing are *smaller* than the costs of behavior; (c) *prescriptive* norm endorses the proposed behavior, and the benefits of behavior for overall wellbeing are *greater* than the costs of behavior; (d) *prescriptive* norm endorses the proposed behavior, and the benefits of behavior for overall wellbeing are *smaller* than the costs of behavior. In the case of the traditional moral dilemma, only one combined situation (proscriptive norm and benefits greater than costs) was included and not the other three combined situations (proscriptive norm and benefits smaller than costs; prescriptive norm and benefits greater than costs; prescriptive norm and benefits smaller than costs).

Second, they used an MPT to depict the mental processes of the agent’s moral judgment (Figure 1).

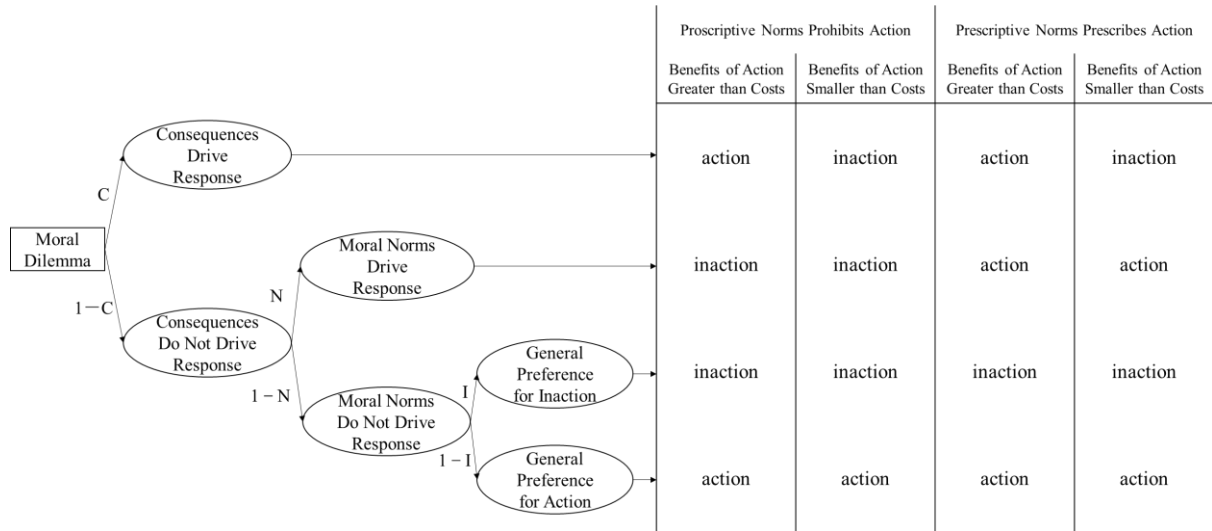


Figure 1. Multinomial processing tree predicting action *versus* inaction responses in moral dilemmas with proscriptive and prescriptive norms and consequences involving benefits of action that are greater or smaller than the costs of action. Retrieved from Gawronski et al. (2017).

Together with the MPT model, the model equations are attached. The sum of probabilities of action and inaction in each dilemma is 1, so we have listed only the equations for action probability. To simplify the equations, let $p(\text{action} \mid \text{proscriptive norm, benefits} > \text{costs})$ be $p1$, let $p(\text{action} \mid \text{proscriptive norm, benefits} < \text{costs})$ be $p2$, let $p(\text{action} \mid \text{prescriptive norm, benefits} > \text{costs})$ be $p3$, let $p(\text{action} \mid \text{prescriptive norm, benefits} < \text{costs})$ be $p4$, and same hereinafter.

$$p1 = C + (1 - C) \times (1 - N) \times (1 - I) \quad (1)$$

$$p2 = (1 - C) \times (1 - N) \times (1 - I) \quad (2)$$

$$p3 = C + (1 - C) \times N + (1 - C) \times (1 - N) \times (1 - I) \quad (3)$$

$$p4 = (1 - C) \times N + (1 - C) \times (1 - N) \times (1 - I) \quad (4)$$

With this model, three parameters could be dissociated using maximum likelihood statistics: Consequences sensitivity (C), Norms sensitivity (N), and generalized Inaction *versus* Action irrespective of consequences and norms (I). Hence, the model was termed the “CNI model.” Gawronski et al. (2017) provided protocols with a MultiTree program (Moshagen,

2010) to generate *C/N/I* parameters (www.bertramgawronski.com/documents/CNI-Model_Materials.zip).

Methodological limitations of the CNI model

The CNI model contributes to the literature because it is claimed to dissociate the three possibilities if the agent makes decisions in a traditional moral dilemma. Therefore, the CNI model can be used to solve several inconsistent findings, such as whether incidental emotions affect moral judgment and how (Gawronski, Conway, Armstrong, Friesdorf, & Hutter, 2018). However, recently Baron and Goodwin demonstrated several theoretical problems underlying the CNI model, such as prohibition of deontological rules (for details, see Baron & Goodwin, 2019). In the present study, we want to highlight some methodological limitations of the CNI model and to solve them using a new algorithm.

First, the CNI model cannot be applied for correlation and regression analyses. *C/N/I* parameters are generated with maximum likelihood statistics, and the parameters are at the group level rather at the individual level. Thus, the CNI model cannot be used in studies aiming to discuss correlations.

Second, the CNI model can only compare the differences between two parameters and one parameter to a specific value. It is inapplicable if multiple comparisons beyond two conditions need to be made.

Lastly (but most importantly), the CNI model hypothesizes the agent first considers whether the consequences of the proposed behavior are beneficial, then, considers whether the proposed behavior is allowed by moral norms, and finally, considers strategies of either generalized action or inaction irrespective of consequences or norms. This *priori* hypothesis is

untenable for two reasons. First, if the agent sequentially considers the decision principles, s/he would not feel dilemmatic when *norms* prohibit action while *consequences* advocate action. The agent will feel dilemmatic only if s/he is simultaneously considering *norms* and *consequences* principles in conflicted situations. Thus, the agent is more likely to simultaneously (rather than sequentially) activate his/her *norms* and *consequences* principles. Second, there could be other sequential processing patterns even if the agent is in a sequential mindset. The sequential processing patterns could be $N \rightarrow C \rightarrow I$ (first consider *norms*, if not, then consider *consequences* and finally, consider strategies of generalized action/inaction), $I \rightarrow C \rightarrow N$ (first obtain a generalized action/inaction preference, then revise it by the *consequences* principle and, finally, revise it by the *norms* principle), $I \rightarrow N \rightarrow C$ (first obtain a generalized action/inaction preference, then revise it by the *norms* principle and, finally, revise it by the *consequences* principle) and other potential sequential response patterns. Taking the $N \rightarrow C \rightarrow I$ pattern as an example, named the “NCI model,” see Figure 2.

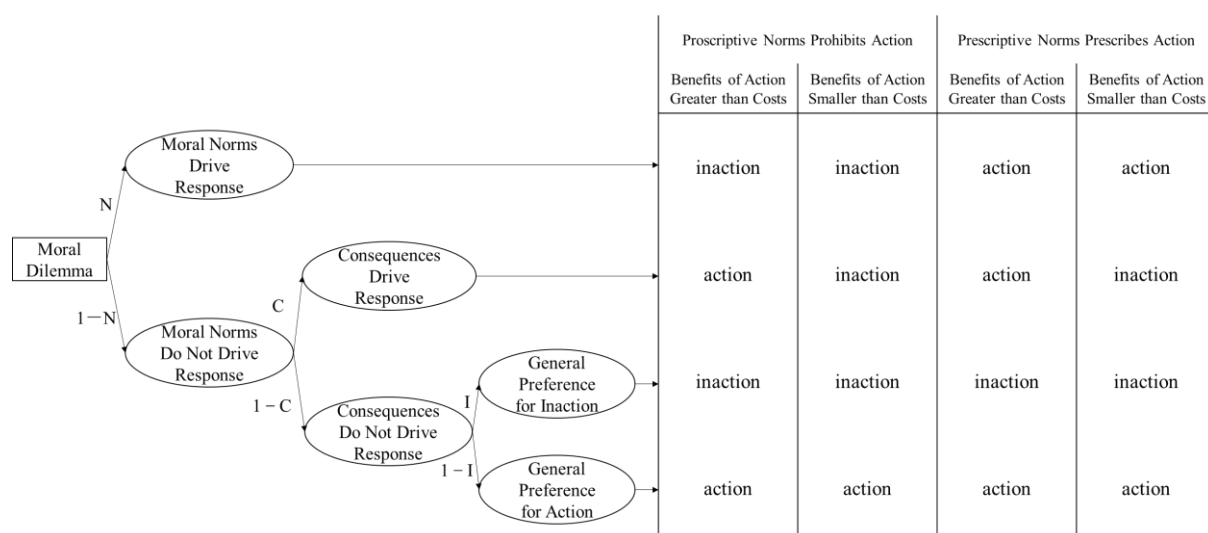


Figure 2. NCI model. The positions of the C parameter and N parameter are exchanged based on the original CNI model.

With the NCI model, we can use model equations to depict the response probabilities of the four combined dilemma situations, too:

$$p1 = (1 - N) \times C + (1 - N) \times (1 - C) \times (1 - I) \quad (5)$$

$$p2 = (1 - N) \times (1 - C) \times (1 - I) \quad (6)$$

$$p3 = N + (1 - N) \times C + (1 - N) \times (1 - C) \times (1 - I) \quad (7)$$

$$p4 = N + (1 - N) \times (1 - C) \times (1 - I) \quad (8)$$

Gawronski et al. (2017) discussed this in their footnote 7: all the reported effects were replicated with the NCI model, and the only differences were that some marginally significant effects in the CNI model became statistically significant with the NCI model. Therefore, they did not discuss further the differences between the NCI model and CNI model. However, if the CNI model and NCI model depicted the observed data equally, then equations (1) to (4) and equations (5) to (8) would be statistically identical to generate the parameters. Taking the N parameter as an example, it can be transformed from equations (1) to (4) so that $N = (-p1 - p2 + p3 + p4)/(2 - p1 + p2 - p3 + p4)$, and also be transformed from equations (5) to (8) into $N = (-p1 - p2 + p3 + p4)/2$. If the CNI model and NCI model are statistically equivalent, these two N parameters should be equal. After conversion, it turns out that $p2 - p1 = p3 - p4$. In the same way, to transform the C parameter based on the equations for the CNI model and NCI model, it turns out that $p1 + p2 = p3 + p4$. Combining these two transformed equations, it would turn out that $p2 = p3$ and $p1 = p4$. These conversions imply that the CNI model and NCI model would generate the same N and C parameters only if $p2 = p3$, and $p1 = p4$. However, this precondition obviously has very low empirical possibility.

The first two limitations of CNI model were due to that the parameters were recorded in group level rather than individual level. The last but the most fatal limitation of CNI model was due to that CNI model presupposed the agent was sequentially rather than parallelly considering the *norms* and *consequences* principles. Given these methodological limitations,

we tried to develop a new algorithm to identify the agent's *norms* and *consequences* sensitivities and overall action/inaction preferences.

CAN algorithm

The traditional moral dilemma is varied into four parallel versions by manipulating the potential moral principles of norms and consequences, and the action is prohibited or advocated by *norms* and *consequences* principles. Thus, we can use a common algebraically subtracting strategy to generate C and N parameters. This strategy is commonly used in the literature, such as Talhelm computed loyalty/nepotism as the amount participants rewarded their friend minus the amount they punished their friend (Talhelm et al., 2014). With respect to the A parameter, we used an aggregate mean strategy to measure the overall action *versus* inaction preferences, which is explained below.

With respect to the C parameter, if individuals are sensitive to consequences, they are more likely to approve the proposal under the conditions of benefits greater than costs than under the conditions of benefits smaller than costs. Therefore, the sensitivity of consequences under proscriptive norms conditions could be represented by $p1 - p2$, and the sensitivity of consequences under prescriptive norms conditions could be represented by $p3 - p4$. Hence, the *consequences* sensitivity is represented by the mean value of the two conditions, i.e., $C = (p1 - p2 + p3 - p4)/2$.

With regard to the N parameter, the sensitivity of norms under conditions of benefits greater than costs could be represented by $p3 - p1$; the sensitivity of norms under conditions of benefits smaller than costs could be represented by $p4 - p2$. Thus, the *norms* sensitivity is represented by the mean value of the two conditions, i.e., $N = (p3 - p1 + p4 - p2)/2$.

For the A parameter, this index is used to represent an individual's overall action/inaction preferences as a whole rather than generalized action/inaction preferences irrespective of norms and consequences. The mean action probability under the four situations could be calculated, i.e., $A = (p_1 + p_2 + p_3 + p_4)/4$. We do not think that the I parameter makes sense under sequential processing in the CNI model, and this is discussed below.

If the C/N parameter is greater (less) than 0, then the participants are identified as being sensitive to supporting (opposing) the *norms/consequences*. The larger the C/N parameter, the more sensitive it is to supporting the *norms/consequences*. If the C/N parameter is not significantly different to 0, then the participants are identified as not being sensitive to *norms/consequences*. The larger the A parameter, the more overall is endorsement of the behavior proposal. If the A parameter is greater (less) than 0.5, then the participants are identified as having an overall action (inaction) preference. If the A parameter is not significantly different to 0.5 while at least one of the C/N parameters is significantly different to 0, the participants are identified as having a pure morality attitude of utilitarian or deontological. If the A parameter is not significantly different to 0.5 while neither the C parameter nor the N parameter is significantly different to 0, the participants are identified as answering randomly.

To differentiate it from the CNI model, we named this new algorithm as "CAN". To demonstrate further the reasonability of the CAN algorithm and the differences between the CAN algorithm and CNI model, the equations of these two approaches will be discussed.

Contrasts in parameters between the CNI model and CAN algorithm

The equations for the C parameter are identical under the two methods. With equations for the CNI model, we can transform them based on equations (1) and (2) into $C = p1 - p2$, and transform them based on equations (3) and (4) into $C = p3 - p4$. On average, $C = (p1 - p2 + p3 - p4)/2$. This equation is identical to the equation under the CAN algorithm. Therefore, we predict that in each study of Gawronski et al. (2017), the mean value of the C parameter under the CNI model will be almost equal to the mean values of C parameters under the CAN algorithm. Given that the C parameter under the CNI model was computed with maximum likelihood statistics at the group level whereas the C parameter under the CAN algorithm was computed with an algebraically subtracting strategy at the individual level, the mean values of these two parameters should be approximately equal rather than absolutely equal.

For the N parameter, we can transform equations (1) to (4) of the CNI model into $N = (p3 - p1 + p4 - p2)/(2 \times (1 - C))$. However, the equation for the N parameter in the CAN algorithm is $N = (p3 - p1 + p4 - p2)/2$. That is, the N parameter under the CAN algorithm divided by $(1 - C)$ will be approximately equal to the N parameter under the CNI model. This is because the CNI model hypothesizes that the agent would consider the *norms* on the basis of not considering *consequences*. This precondition is untenable because it is entirely possible for the agent to consider *norms* first. Then, the sequential processing model should be the NCI model. If so, we can transform equations (5) to (8) of the NCI model into $N = (p3 - p1 + p4 - p2)/2$, which is identical to the CAN algorithm. Thus, we predict that $N/(1 - C)$ under the CAN algorithm will be approximately equal to the N parameter value under the CNI model.

With regard to I and A parameters, in the logic of the CNI model, the agent will consider the generalized action/inaction preference based on not considering *norms* and *consequences*.

Similar to the N parameter, this precondition is not reasonable. The agent can have a generalized action/inaction preference to the behavior proposal first, and then her/his choices will be influenced by *norms* and *consequences* principles so that the choices are corrected based on the corresponding principles. Thus, the I parameter under the CNI model is not credible. We gave up the endeavor to identify the generalized inaction/action preferences irrespective of norms and consequences. Instead, the overall tendency of the agent's action/inaction is more critical to reflect the overall preferences. If the agent makes decisions purely according to *norms* and *consequences* principles, p_2 would tend to be 0 and p_3 would tend to be 1 because norms and consequences principles prohibit or advocate action; also the mean value of p_1 and p_4 would tend to be 0.5 because the *norms* principle and *consequences* principle are conflicted in terms of prohibiting or advocating action. Overall, the A parameter (i.e., $(p_1 + p_2 + p_3 + p_4)/4$) should have no differences to 0.5 if the agent makes decisions based purely on norms and consequences principles, or the agent is just answering randomly. Thus, the A parameter can represent the agent's overall action/inaction preferences.

Overview of the present study

Based on the methodological discussion above, we can make certain predictions. The C parameter under the CNI model will be approximately equal to the C parameter under the CAN algorithm ($H1$). The N parameter under the CNI model will be approximately equal to $N/(1 - C)$ under the CAN algorithm ($H2$). The I parameter under the CNI model represents the agent's generalized *inaction* versus *action* preferences, whereas the A parameter under the CAN algorithm represents the agent's overall *action* versus *inaction* preferences. Thus, these two parameters will be reversed around 0.5. If the I parameter is higher than 0.5, the A parameter

will be lower than 0.5, or *vice versa* (*H3*). Although the CNI model and CAN algorithm are algebraically different, the bias might be balanced systematically across between-subject conditions. Consequently, the tests of between-subject differences of the parameters generated from the CNI model and CAN algorithm could be almost identical (*H4*).

METHODS

In order to test the predictions, we re-analyzed the raw data of Gawronski et al. (2017) in which the CNI model was proposed and tested.

First, we downloaded the raw data of Gawronski et al. (2017) from <https://osf.io/xt66w/>. Then, we re-analyzed the raw data with the CNI model to ensure that the results of Gawronski et al. (2017) were reproducible. Our re-analysis results were identical to the results they reported.

Second, we used the CAN algorithm to generate $C/N/A$ parameters, and also calculated $N/(1 - C)$ with the mean values of C and N parameters. After that, we tested the hypotheses stated above.

Finally, because the $C/N/A$ parameters generated from the CAN algorithm are at the individual level, these parameters could be used for correlation and other analyses. Thus, we tried to make the Pearson correlation analysis between psychopathy scale rating and parameters in Gawronski et al. (2017)'s Study 4.

RESULTS

With the CNI model, Gawronski et al. (2017) conducted four formal studies and one supplemental study. Each study was replicated based on recent concerns about the reproducibility of psychological findings (Open Science Collaboration, 2015). Thus, there were

10 studies in total. Study 1a/b discussed the gender differences in moral decision-making because it remained ambiguous in the traditional-dilemma approach (see Friesdorf, Conway, & Gawronski, 2015). Study 2a/b explored the effects of cognitive load on moral decision-making (see Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008). Study 3a/b was on the effect of question-framing because one study demonstrated that personal force enhanced deontological responses (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Study 4a/b explored the relationship between subclinical psychopathy level and utilitarian responses (Bartels & Pizarro, 2011; Kahane, Everett, Earp, Farias, & Savulescu, 2015). Study S1a/b discussed the effects of harm salience on moral decision-making (see the relevant study, Conway & Gawronski, 2013).

Gawronski et al. (2017) conducted analyses of a traditional dilemma, process dissociation (Conway & Gawronski, 2013) and the CNI model for each study. In our re-analysis, we replicated only the analysis of the CNI model and re-analyzed the raw data with the CAN algorithm. The patterns of re-analysis results were almost identical across all studies. Hence, we present the results of Study 1a/b; the remaining results are in the Appendices.

Test the hypotheses of present study

Re-analyses of Gawronski et al. (2017)'s Study 1a

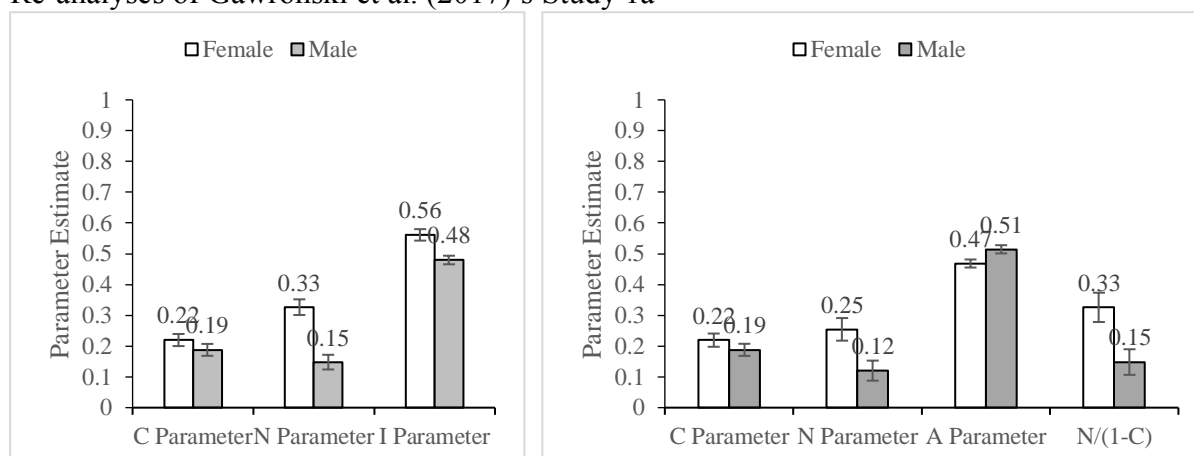


Figure 3. Results of Gawronski et al. (2017)'s Study 1a obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table 1. Test of gender differences with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 1a

Parameters		Results	Conclusion contrast	Results	Parameters
CNI Model	C	$\Delta G^2(1) = 1.34, p = .247, d = 0.164$	identical	$t(199) = 1.08, p = .281, d = 0.153$	C
	N	$\Delta G^2(1) = 26.00, p < .001, d = 0.726$	identical	$t(199) = 2.74, p = .007, d = 0.387$	N
	I	$\Delta G^2(1) = 12.34, p < .001, d = 0.504$	identical	$t(199) = 2.45, p = .015, d = 0.346$	A

Re-analyses of Gawronski et al. (2017)'s Study 1b

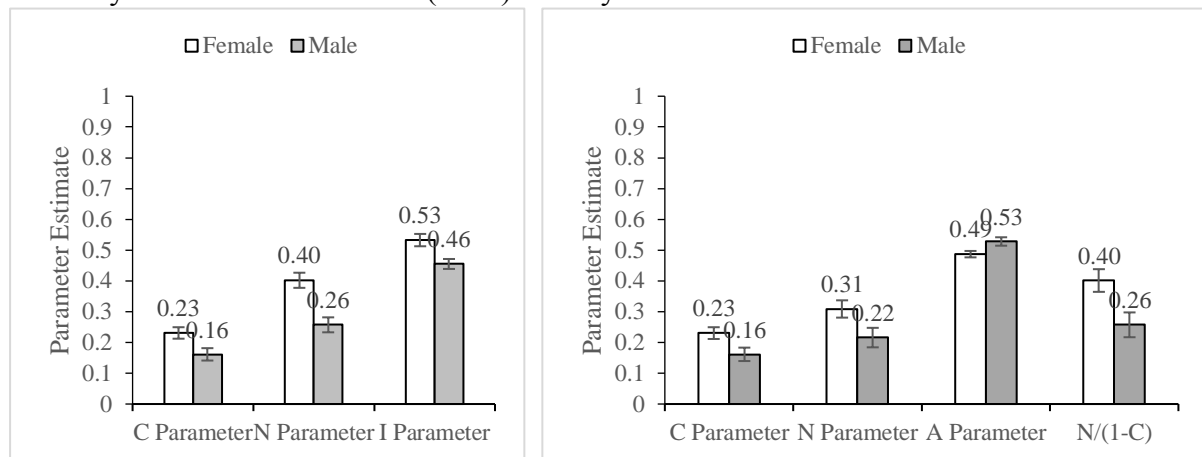


Figure 4. Results of Gawronski et al. (2017)'s Study 1b obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table 2. Test of gender differences with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 1b

Parameters		Results	Conclusion contrast	Results	Parameters
CNI Model	C	$\Delta G^2(1) = 6.43, p = .011, d = 0.364$	identical	$t(195) = 2.39, p = .018, d = 0.153$	CAN Algorithm
	N	$\Delta G^2(1) = 17.43, p < .001, d = 0.599$	identical	$t(195) = 2.20, p = .029, d = 0.314$	

	$\Delta G^2(1) = 9.12, p$		$t(195) = 2.41, p$	
I	$= .003,$	identical	$= .017, d =$	A
	$d = 0.428$		0.344	

Across the 10 studies, all the predictions were validated, as shown in *Figure 3 & 4*. The C parameter generated from the CNI model and CAN algorithm was approximately equal. The N parameter from the CAN algorithm was slightly smaller than that from the CNI model, and $N/(1 - C)$ in the CAN algorithm was approximately equal to the N parameter in the CNI model. The I parameter under the CNI model and A parameter under the CAN algorithm would be reversed around 0.5 because their statistical implications were different. The larger the I parameter means the more generalized inaction *versus* action preferences, whereas the larger the A parameter denotes the more overall action *versus* inaction tendencies. Furthermore, the differences in parameters across between-subject conditions were almost identical to the CNI model and CAN algorithm, as shown in Table 1 & 2. The independent sample t -test with $C/A/N$ parameters was more stringent than maximum likelihood statistics with $C/N/I$ parameters. Thus, a few marginally significant and low-significant results under the CNI model became non-significant under the CAN algorithm (see the Appendices).

Demonstrating the statistical advantage of CAN algorithm

Moreover, the CAN algorithm could be used in correlation analysis. To demonstrate this statistical advantage over the CNI model, we ran a correlation analysis with the raw data of Study 4. In the latter, participants were divided artificially into “low” and “high” psychopathy conditions based on their scores on a psychopathy scale. We ran a Pearson correlation analysis between their psychopathy scores and $C/A/N$ parameters. In Study 4a, the psychopathy score was not correlated significantly with the C parameter ($r = -0.122, p = .100$), correlated significantly with the N parameter ($r = -0.153, p = .038$), and not correlated

significantly with the A parameter ($r = 0.106, p = .152$). These results are congruent with tests on differences in $C/A/N$ parameters, but incongruent with the test on differences in C/I parameters (see Table A5 in the appendices). In Study 4b, the psychopathy score was correlated significantly with the C parameter ($r = -0.307, p < .001$) and N parameter ($r = -0.394, p < .001$) but not correlated significantly with the A parameter ($r = 0.103, p = .149$). These results also supported the test results on differences based on the CAN algorithm, but did not support the test results on differences in the I parameter (Table A6 in the Appendices).

Overall, the correlation analyses supported the results obtained from the differences test with $C/A/N$ parameters while a little varied comparing to the differences test with $C/N/I$ parameters. Methodologically, CAN algorithm generated the parameters in individual level so that the parameters could be used for correlation and other common analyses, which is an advantage over the CNI model.

DISCUSSIONS

All the hypotheses were verified, and we ran another correlation analysis to demonstrate that the parameters obtained from the CAN algorithm could be used for correlation analyses (which is an advantage over the CNI model). The $C/A/N$ parameters are grounded at the individual level, so that they can be used for much more conventional analyses in a wide range of research designs.

Conception manipulation development

The CAN algorithm is based on the CNI model in theoretical aspects. Originally, the traditional dilemma only considered a scenario in which proscriptive norms and benefits were greater than costs, such as the trolley-car paradigm (Foot, 1967) and footbridge paradigm

(Thomson, 1976). Decades later, Conway and Gawronski (2013) explored two types of scenarios: consistent edition (proscriptive norms and benefits smaller than costs) and inconsistent edition (proscriptive norms and benefits greater than costs). Gawronski et al. (2017) varied the norms and consequences underlying the dilemma, and discussed four types of scenarios: proscriptive/prescriptive norm with benefits greater/smaller than costs. The development of conception manipulation deepened insights on moral decision-making.

However, criticisms remain about the four-edition conception manipulation. Baron and Goodwin (2019) queried that, among the four editions of the same dilemma, the norms underlying the proscriptive editions and the prescriptive editions might not be the same moral norm. For example, in the dilemma of transplant, the norm of proscriptive editions is that *we should not harm other people*, whereas the norm of prescriptive editions is that *we should stop someone doing something harmful to others*. These two situations are essentially different on moral norms.

Nevertheless, we think the four-edition conception manipulation contributed meaningfully. The conception manipulation is event-oriented but not moral principle-oriented. In the four editions of each dilemma, the event is consistent across editions but the underlying norms and consequences vary. Whether the agent is *sensitive* when the underlying norms and consequences change on the same event is important. Moreover, proscriptive and prescriptive morality are very important in people's daily lives: they are the two facets of moral regulation (Janoff-Bulman, Sheikh, & Hepp, 2009). Thus, we agree that proscriptive scenarios and prescriptive scenarios have different norms, but we do not think it matters in measurement of norm *sensitivity*.

The methodological development of CAN algorithm

The CAN algorithm reserved the theoretical conception manipulation development and also fixed the methodological limitations of the CNI model. The CNI model cannot be used for correlation analysis or multiple comparisons because the data obtained by the CNI model are represented at the group level rather than at the individual level. The CAN algorithm algebraically generates the parameters, and parameters data are represented at the individual level. Moreover, the CNI model is suitable only for binary categoric data, whereas the CAN algorithm can also be applied in continuous-scale data. Therefore, $C/A/N$ parameters can be used in a wide scope of research designs and data analyses.

The most serious methodologic limitation is that the CNI model presupposes that the agent sequentially considers consequences, norms, and generalized inaction/action preferences irrespective of norms and consequences. This precondition is questionable and leads to the N parameter being overestimated artificially. As the data re-analysis demonstrated, the N parameter under the CNI model approximately equaled $N/(1 - C)$ under the CAN algorithm. The value of $(1 - C)$ under the CAN algorithm was $[0, 1]$, so the N parameter under the CNI model was systematically larger than the N parameter under the CAN algorithm. The sequential process presuppose of CNI model also makes I parameter dubious as it is claimed to depict the extent of the agent's generalized inaction/action preferences on the basis of not considering *consequences* or *norms*. Therefore, the CAN algorithm adopted a commonly used subtracting strategy to generate C and N parameters and setup an overall action *versus* inaction preferences index: the A parameter.

Even though the agent processes moral decision-making sequentially, the NCI model is more credible than the CNI model. As demonstrated by the theoretical model of social intuition, people react emotionally first and revise their decisions cognitively later. People consider norms intuitively at first and consider consequences rationally later (Haidt, 2001). Therefore, the NCI model is more reasonable even if the style of the sequential process makes sense. However, increasing evidence implies that emotional and cognitive processes are parallel and independent rather than sequential (Cushman, Young, & Greene, 2010; Greene, 2009; Hutcherson, Montaser-Kouhsari, Woodward, & Rangel, 2015; Paxton & Greene, 2010). Thus, the CAN algorithm is more appropriate for demonstration of people's moral preferences, especially C and N parameters.

A more reasonable multinomial process tree model also supports CAN algorithm

Processing tree models are powerful frameworks to discuss potentially conflicted cognitive processes (Calanchini, Rivers, Klauer, & Sherman, 2018; Hutter & Klauer, 2016). Hence, an alternative MPT was constructed based on previous theoretical and empirical evidences (*Figure 5* and named the "DNA model". Bago and De Neys proposed a corrective dual-process model of moral cognition because they found that participants were intuitively utilitarian with a two-response paradigm. Furthermore, they concluded that the agent's final moral judgment was dependent upon the absolute and relative strength between competing deontological and utilitarian intuitions (Bago & De Neys, 2019). Thus, in the MPT, we hypothesized that the driving forces from moral norms and consequences were parallel, and that the response pattern was dependent upon which driving force was stronger.

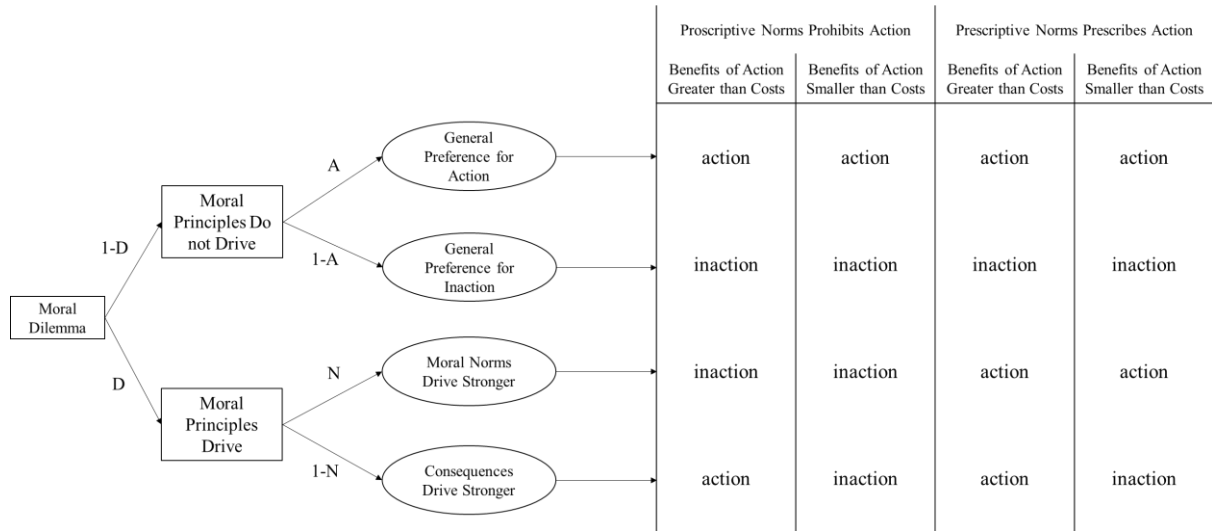


Figure 5. An alternative multinomial processing tree model constructed on the corrective dual-process model of morality. D denotes that the agent's choices are Driven by moral principles (Norms or Consequences), N denotes the agent's choices are driven stronger by Norms. $(1 - N)$ denotes the agent's choices are driven stronger by Consequences. A denotes the agent's choices are driven by an overall preference for action.

Together with the DNA model, four equations can be constructed:

$$p1 = (1 - D) \times A + D \times (1 - N) \quad (9)$$

$$p2 = (1 - D) \times A \quad (10)$$

$$p3 = (1 - D) \times A + D \times N + D \times (1 - N) \quad (11)$$

$$p4 = (1 - D) \times A + D \times N \quad (12)$$

In the DNA model, the consequences sensitivity and norms sensitivity can also be calculated algebraically. Consequences sensitivity can be represented by $D \times (1 - N)$ because the agent should be at first sensitive to moral principles, and then to the Consequences principle. The algebraic expression of $D \times (1 - N)$ can be represented by $(p1 - p2 + p3 - p4)/2$ based on equations (9) to (12). Similarly, norms sensitivity can be represented by $D \times N$ and, further transformed from equations (9) to (12), it would be $(p3 - p1 + p4 - p2)/2$. These two indices were exactly identical to the CAN algorithm. Actually, the DNA model more approximately depicts the moral decision-making process according to the literature (Bago & De Neys, 2019; Neys & Pennycook, 2019).

As to the I parameter in the CNI model, it is problematic if we re-inspect it from the perspective of the DNA model. In the DNA model, the general preference for action irrespective of moral principles could be described as $((1 - D) \times A)$, and its probability is depicted as p_2 according to the equations. p_2 portrays the probability when the agent endorses the behavioral proposal which is prohibited by both norms and consequences principles. Therefore, $((1 - D) \times A)$ is more in line with the implication that the agent endorses the behavioral proposal irrespective of norms and consequences. Similarly, the general preference for inaction irrespective of moral principles could be described as $((1 - D) \times (1 - A))$, and its probability is depicted as $(1 - p_3)$. $(1 - p_3)$ portrays the probability when the agent declines the behavioral proposal which is advocated by both norms and consequences principles. Therefore, $((1 - D) \times (1 - A))$ is more in line with the implication that the agent declines the behavioral proposal irrespective of norms and consequences. Thus, in the DNA model, the generalized action preference irrespective of norms and consequences is p_2 , while the generalized inaction preference irrespective of norms and consequences is $1 - p_3$. As hypothesized by Gawronski et al. (2017), the sum of these two preference probabilities is 1. If that, $p_2 + (1 - p_3) = 1$, in turn, $p_2 = p_3$. It means that only when $p_2 = p_3$, the sum of the probabilities of generalized action and inaction preferences could be 1. However, it is obviously of little possibility. Hence that, the I parameter in the CNI model claimed to depict the generalized inaction/action preferences irrespective of norms and consequences is not credible. On the contrary, the A parameter of the CAN algorithm could be easily understood and credible in methodological connotation as it depicts the *overall* action/inaction preferences among the four editions of moral dilemma.

CONCLUSION

In summary, we addressed the methodological limitations of the CNI model and fixed these limitations with a new algorithm: the CAN. The CNI model presupposes that the agent sequentially considers consequences, norms, and generalized inaction/action preferences irrespective of consequences and norms in his/her moral decision-making process. We provided theoretical evidence that the decision-making process is more likely to be parallel with norms and consequences, and developed the CAN algorithm. The CNI model generated the parameters at the group level, and we calculated the parameters algebraically at the individual level so that the CAN algorithm was suitable for a larger range of research designs and conventional statistical analyses.

REFERENCES

- Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, 148(10), 1782-1801.
- Baron, J., & Goodwin, G. P. (2019). Consequences, Norms, and Inaction: A Comment. SSRN Electronic Journal. doi:10.2139/ssrn.3378106
- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, 121(1), 154-161.
- Bentham, J. (1996). *An Introduction to the Principles of Morals and Legislation*. New York: Oxford University Press, USA. (Original work published 1781).
- Calanchini, J., Rivers, A. M., Klauer, K. C., & Sherman, J. W. (2018). Chapter Two - Multinomial processing trees as theoretical bridges between cognitive and social psychology. In K. D. Federmeier (Ed.), *Psychology of Learning and Motivation*, pp. 39-65): Academic Press.
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of Personality and Social Psychology*, 104(2), 216-235.
- Cushman, F., Young, L., & Greene, J. D. (2010). Multi-system moral psychology. In J. D. e. al. (Ed.), *The Moral Psychology Handbook* (pp. 47-71). New York: Oxford University Press.
- Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 2(2), 152-161.

- Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: a process dissociation analysis. *Personality and Social Psychology Bulletin*, 41(5), 696-713.
- Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hutter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology*, 113(3), 343-376.
- Gawronski, B., & Beer, J. S. (2017). What makes moral dilemma judgments "utilitarian" or "deontological"? *Social Neuroscience*, 12(6), 626-632.
- Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hutter, M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion*, 18(7), 989-1008.
- Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, 45(3), 581-584.
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107(3), 1144-1154.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814-834.

- Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and Utilitarian Appraisals of Moral Dilemmas Are Encoded in Separate Areas and Integrated in Ventromedial Prefrontal Cortex. *Journal of Neuroscience*, 35(36), 12593-12605.
- Hutter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, 27(1), 116-159.
- Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: two faces of moral regulation. *Journal of Personality and Social Psychology*, 96(3), 521-537.
- Kahane, G., Everett, J. A., 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.
- Kant, I., & Gregor, M. J. (1997). *Groundwork of the metaphysics of morals*. Cambridge, UK: Cambridge University Press.
- Mill, J. S. (1872). *The Logic of the Moral Sciences*. Chicago, Illinois, US: Open Court.
- Moshagen, M. (2010). multiTree: a computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42(1), 42-54.
- Neys, W. D., & Pennycook, G. (2019). Logic, Fast and Slow: Advances in Dual-Process Theorizing. *Current Directions in Psychological Science*, 28(5), 503-509.
- Open Science, Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Paxton, J. M., & Greene, J. D. (2010). Moral reasoning: hints and allegations. *Topics in*

Cognitive Science, 2(3), 511-527.

Rawls, J. R. (1971). *A Theory of Justice*. Cambridge, Massachusetts, USA; London, England: The Belknap Press of Harvard University Press.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture. *Science*, 344(6184), 603-608.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *Monist*, 59(2), 204-217.

APPENDICES

Re-analyses of Gawronski et al. (2017)'s Study 2a

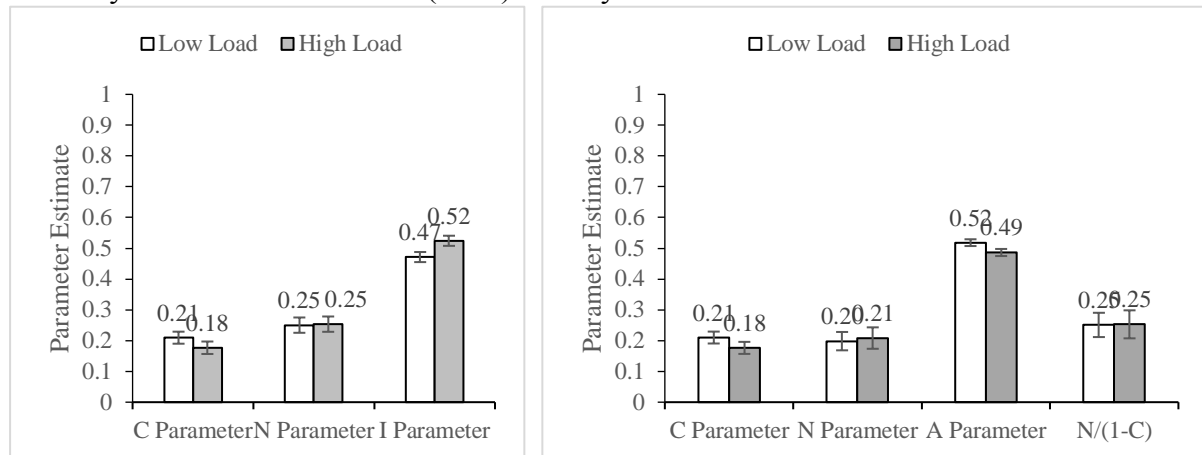


Figure A1. Results of Gawronski et al. (2017)'s Study 2a obtained from CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table A1. Test of differences in cognitive load with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 2a

Parameters		Results	Conclusion contrast	Results	Parameters
CNI Model	C	$\Delta G^2(1) = 1.35, p = .245, d = 0.168$	identical	$t(192) = 1.22, p = .223, d = 0.175$	CAN Algorithm
	N	$\Delta G^2(1) = 0.01, p = .927, d = 0.013$	identical	$t(192) = 0.22, p = .827, d = 0.031$	
	I	$\Delta G^2(1) = 5.19, p = .023, d = 0.328$	identical	$t(192) = 2.01, p = .045, d = 0.289$	

Re-analyses of Gawronski et al. (2017)'s Study 2b

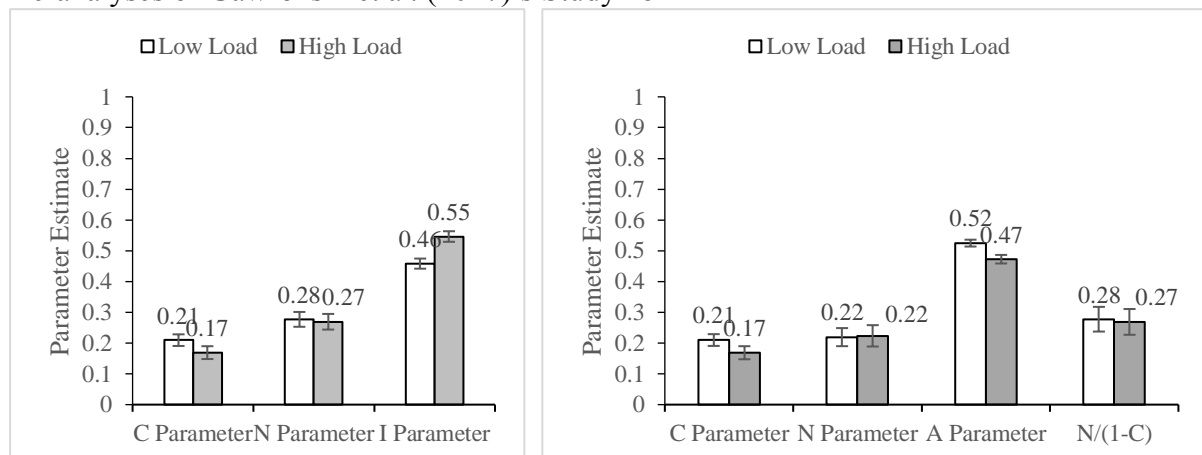


Figure A2. Results of Gawronski et al. (2017)'s Study 2b obtained from the CNI model (left) and CAN

algorithm (right). Error bars represent ± 1 SE.

Table A2. Test of differences in cognitive load with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 2b

Parameters	Results	Conclusion contrast	Results	Parameters
CNI Model	$\Delta G^2(1) = 2.08, p = .149, d = 0.209$	identical	$t(192) = 1.45, p = .149, d = 0.209$	C
	$\Delta G^2(1) = 0.05, p = .826, d = 0.032$	identical	$t(192) = 0.09, p = .927, d = 0.013$	N
	$\Delta G^2(1) = 13.77, p < .001, d = 0.535$	identical	$t(192) = 2.97, p = .003, d = 0.428$	A

CAN
Algorithm

Re-analyses of Gawronski et al. (2017)'s Study 3a

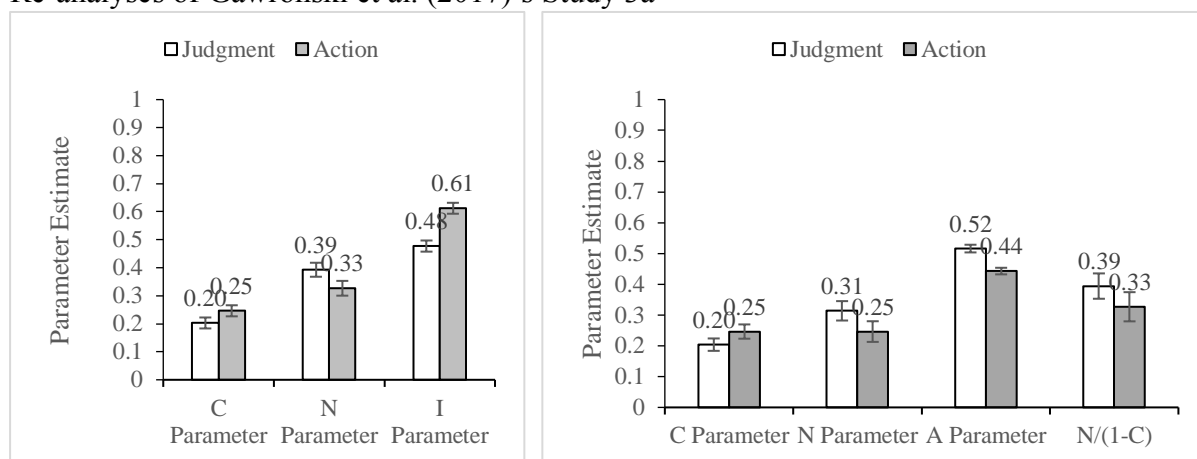


Figure A3. Results of Gawronski et al. (2017)'s Study 3a obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table A3. Tests of differences in question-framing with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 3a

Parameters	Results	Conclusion contrast	Results	Parameters
CNI Model	$\Delta G^2(1) = 2.44, p = .118, d = 0.230$	identical	$t(184) = 1.39, p = .168, d = 0.203$	C
	$\Delta G^2(1) = 3.31, p = .069, d = 0.268$	basically identical	$t(184) = 1.47, p = .144, d = 0.013$	N
	$\Delta G^2(1) = 35.18, p < .001, d = 0.713$	identical	$t(184) = 4.43, p < .001, d = 0.650$	A

CAN
Algorithm

Re-analyses of Gawronski et al. (2017)'s Study 3b

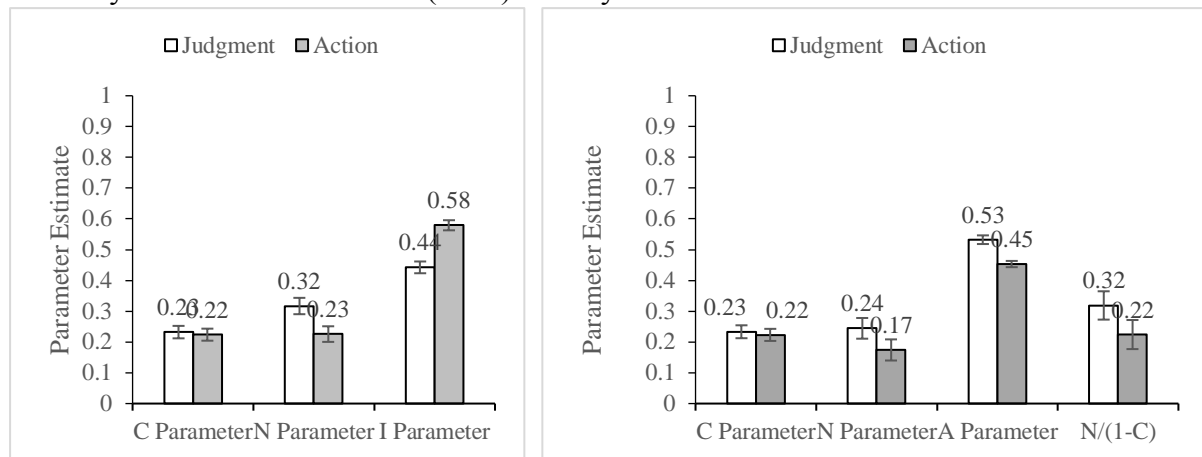


Figure A4. Results of Gawronski et al. (2017)'s Study 3b obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table A4. Test of differences in question-framing with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 3b

Parameters		Results	Conclusion contrast	Results	Parameters
CNI Model	<i>C</i>	$\Delta G^2(1) = 0.09, p = .767, d = 0.043$	identical	$t(187) = 0.36, p = .721, d = 0.052$	<i>C</i>
	<i>N</i>	$\Delta G^2(1) = 6.15, p = .013, d = 0.363$	discrepant	$t(187) = 1.46, p = .147, d = 0.212$	<i>N</i>
	<i>I</i>	$\Delta G^2(1) = 29.50, p < .001, d = 0.799$	identical	$t(187) = 4.62, p < .001, d = 0.673$	<i>A</i>
					CAN Algorithm

Re-analyses of Gawronski et al. (2017)'s Study 4a

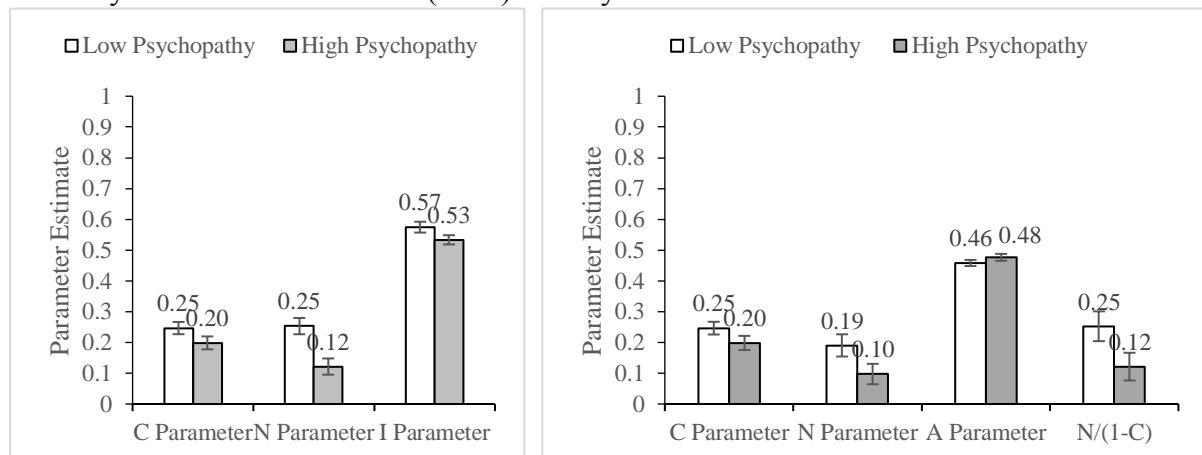


Figure A5. Results of Gawronski et al. (2017)'s Study 4a obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table A5. Test of psychopathy differences with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 4a

Parameters		Results	Conclusion contrast	Results	Parameters	
CNI Model	<i>C</i>	$\Delta G^2(1) = 2.77, p = .096, d = 0.247$	discrepant	$t(182) = 1.56, p = .121, d = 0.230$	<i>C</i>	CAN Algorithm
	<i>N</i>	$\Delta G^2(1) = 12.35, p < .001, d = 0.521$	basically identical	$t(182) = 1.89, p = .060, d = 0.279$	<i>N</i>	
	<i>I</i>	$\Delta G^2(1) = 3.15, p = .076, d = 0.262$	discrepant	$t(182) = 1.22, p = .223, d = 0.180$	<i>A</i>	

Re-analyses of Gawronski et al. (2017)'s Study 4b

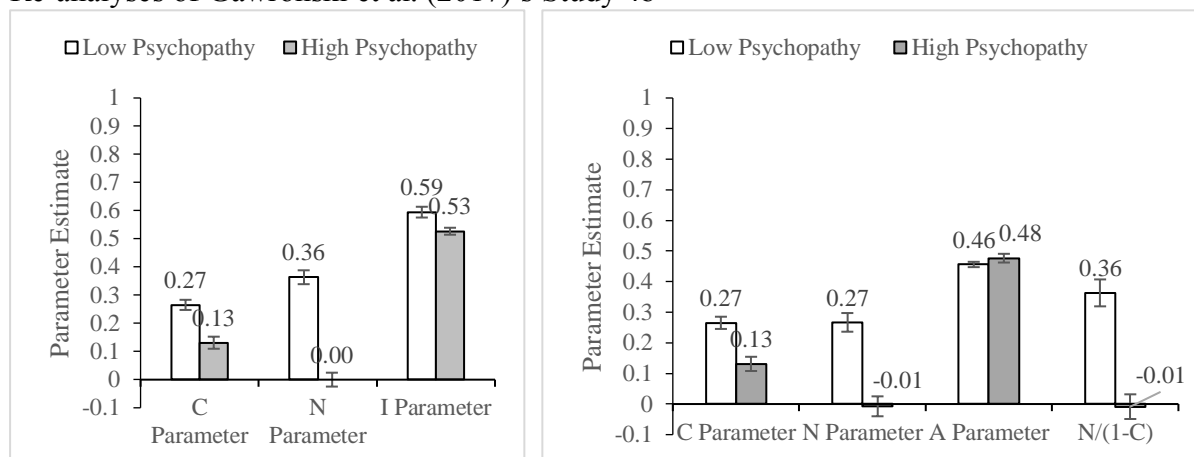
Figure A6. Results of Gawronski et al. (2017)'s Study 4b obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table A6. Test of psychopathy differences with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study 4b

Parameters		Results	Conclusion contrast	Results	Parameters	
CNI Model	<i>C</i>	$\Delta G^2(1) = 23.13, p < .001, d = 0.695$	identical	$t(196) = 4.40, p < .001, d = 0.629$	<i>C</i>	CAN Algorithm
	<i>N</i>	$\Delta G^2(1) = 111.80, p < .001, d = 1.48$	identical	$t(196) = 6.12, p < .001, d = 0.875$	<i>N</i>	
	<i>I</i>	$\Delta G^2(1) = 8.90, p = .003, d = 0.406$	discrepant	$t(196) = 1.28, p = .202^a, d = 0.183$	<i>A</i>	

^aLevene's test was significant ($p < 0.05$), suggesting violation of the equal-variance assumption.

Re-analyses of Gawronski et al. (2017)'s Study S1a

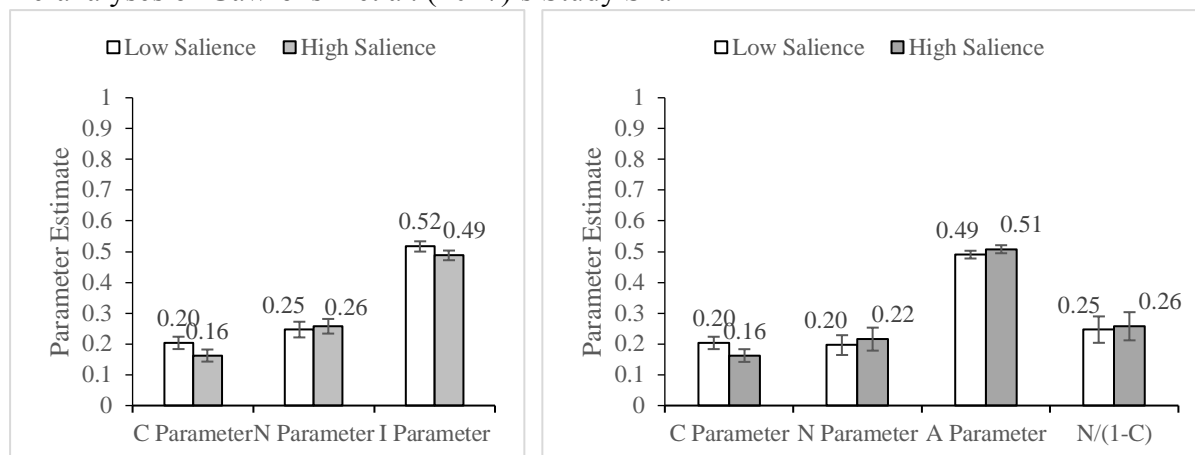


Figure A7. Results of Gawronski et al. (2017)'s Study S1a obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table A7. Test of differences in harm salience with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study S1a

Parameters		Results	Conclusion contrast	Results	Parameters
CNI Model	<i>C</i>	$\Delta G^2(1) = 2.15, p = .143, d = 0.211$	identical	$t(193) = 1.42, p = .156, d = 0.204$	<i>C</i>
	<i>N</i>	$\Delta G^2(1) = 0.10, p = .758, d = 0.044$	identical	$t(193) = 0.39, p = .696, d = 0.056$	<i>N</i>
	<i>I</i>	$\Delta G^2(1) = 1.60, p = .206, d = 0.182$	identical	$t(193) = 0.97, p = .334, d = 0.139$	<i>A</i>
					CAN Algorithm

Re-analyses of Gawronski et al. (2017)'s Study S1b

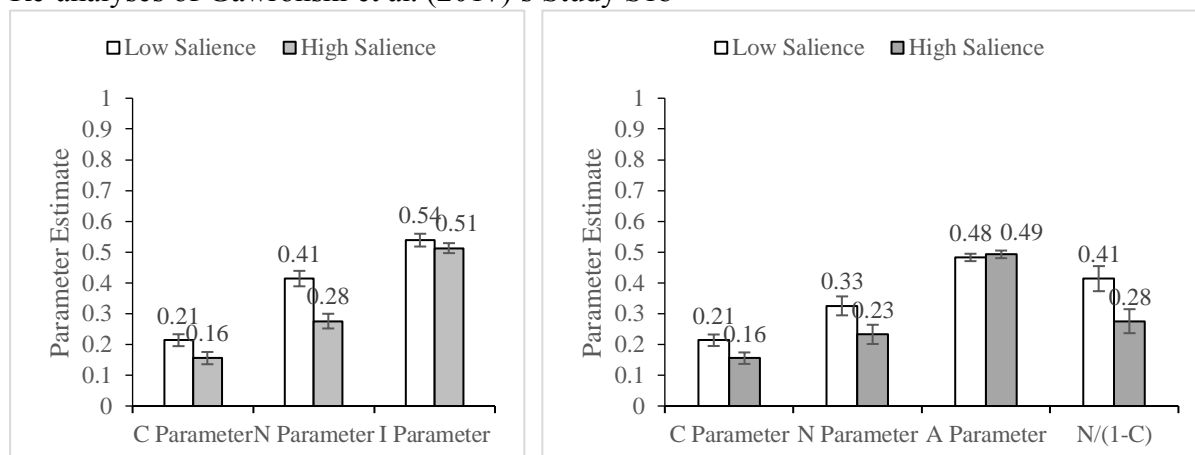


Figure A8. Results of Gawronski et al. (2017)'s Study S1b obtained from the CNI model (left) and CAN algorithm (right). Error bars represent ± 1 SE.

Table A8. Test of differences in harm salience with the CNI model and CAN algorithm in Gawronski et al. (2017)'s Study S1b

Parameters		Results	Conclusion contrast	Results	Parameters	
CNI Model	<i>C</i>	$\Delta G^2(1) = 4.40, p = .036, d = 0.305$	identical	$t(189) = 2.20, p = .029, d = 0.204$	<i>C</i>	CAN Algorithm
	<i>N</i>	$\Delta G^2(1) = 15.79, p < .001, d = 0.580$	identical	$t(189) = 2.11, p = .036, d = 0.305$	<i>N</i>	
	<i>I</i>	$\Delta G^2(1) = 0.97, p = .325, d = 0.144$	identical	$t(189) = 0.58, p = .563, d = 0.084$	<i>A</i>	