stats

MDPI

Article

Scoring Individual Moral Inclination for the CNI Test

Yi Chen *, Benjamin Lugu, Wenchao Ma and Hyemin Han *

Department of Educational Studies in Psychology, Research Methodology and Counseling, The University of Alabama, Tuscaloosa, AL 35401, USA; bklugu@crimson.ua.edu (B.L.); wenchao.ma@ua.edu (W.M.)
* Correspondence: ychen195@crimson.ua.edu (Y.C.); hyemin.han@ua.edu (H.H.)

Abstract: Item response theory (IRT) is a modern psychometric framework for estimating respondents' latent traits (e.g., ability, attitude, and personality) based on their responses to a set of questions in psychological tests. The current study adopted an item response tree (IRTree) method, which combines the tree model with IRT models for handling the sequential process of responding to a test item, to score individual moral inclination for the CNI test—a broadly adopted model for examining humans' moral decision-making with three parameters generated: sensitivity to moral norms, sensitivity to consequences, and inaction preference. Compared to previous models for the CNI test, the resulting EIRTree-CNI Model is able to generate individual scores without increasing the number of items (thus, less subject fatigue or compromised response quality) or employing a post hoc approach that is deemed statistically suboptimal. The model fits the data well, and the subsequent test also supported the concurrent validity and the predictive validity of the model. Limitations are discussed further.

Keywords: moral dilemma judgment; CNI model; item response theory

1. Introduction

1.1. Item Response Theory

Item response theory (IRT) is a modern psychometric framework for estimating respondents' latent traits (e.g., ability, attitude, and personality) based on their responses to a set of questions in psychological tests. There are many different IRT models, each considering different facets of item characteristics (e.g., [1]). For example, the responses to a multiple-choice item in a math test can be graded, yielding a binary variable $Y = 1$ if the response is correct and $Y = 0$ if the response if not correct. The one-parameter logistic (1PL) IRT model [2] or Rasch model [3] assumes that the probability of answering this item correctly is a function of the respondent's math ability θ and this item's difficulty β, as defined in

$$P(Y = 1|\theta,\ \beta) = \frac{1}{1 + \exp(\theta - \beta)}$$

IRT models allow us to estimate all respondents' ability based on their responses and also quantify the items' characteristics, such as item difficulty.

In this paper, we adopted an item response tree (IRTree) method, which combines the tree model with IRT models for handling the sequential process of responding to a test item [4]. The IRTree models have gained increasing attention in psychology and social sciences. For example, they have been used to identify and accommodate different response styles [5], model answer change behaviors [6], and analyze multiple strategies used in problem solving [7,8]. Traditional IRTree models usually involve a single path to each observed score, but this study adopts an extended version of the IRTree method, similar to [8], allowing for two or more paths to the same observed options. This is referred to as the EIRTree model in this paper.

check for updates

Citation: Chen, Y.; Lugu, B.; Ma, W.; Han, H. Scoring Individual Moral Inclination for the CNI Test. *Stats* **2024**, *7*, 894–905. https://doi.org/10.3390/stats7030054

Academic Editor: Wei Zhu

Received: 1 July 2024
Revised: 18 August 2024
Accepted: 20 August 2024
Published: 23 August 2024

1.2. Moral Dilemmas and CNI Model

Moral dilemmas have a long history in moral philosophy. They have been used as an effective tool to test our moral intuition and reasoning, and establish principles for addressing difficult moral issues in the real world [9]. Think of this scenario:

You are a church official during the Nazi occupation of the Netherlands. You have an official agreement with the occupying forces that protects anyone under your care. You publicly spoke on behalf of a Jewish family, annoying the Nazis. To demonstrate their power, the Nazis demand that you kill the members of the Jewish family. Otherwise, they will execute dozens of Dutch people they have incarcerated for political reasons. Is it acceptable in this case to kill the Jewish family [10]?

When faced with such a problem, responders usually hesitate to give an answer [11], or struggle with the rationale of their decision [12]. In moral philosophy, an answer of "Yes" to the above question would be classified as utilitarian in nature, which means adopting an action that promotes the greatest good (saving more people). The utilitarian principle holds that moral action is right and only right if it leads to the greatest good for the greatest number of people [13]. It reflects how sensitive the individual is to the consequence of the action. Whereas "No" would be considered a deontological response, which means refusing to choose a morally wrong action—killing. Deontology deems action as more important than consequence [14]: certain rights or duties (e.g., do not harm) must be respected in all scenarios. In other words, the deontological preference indicates norm sensitivity.

These moral dilemmas were extensively adopted as a research instrument by moral psychologists subsequent to the work of [15], who designed a series of such dilemmas to investigate the neural underpinnings of and the dilemma characteristic's influence on the role of reason and emotion in moral judgment. However, one drawback of this pervasively utilized measurement is that all dilemmas here apply only proscriptive moral norms (that is, if participants make a decision based on the moral norm, they should always answer "no"), and the benefits of action always outweigh its drawbacks (answering "yes" will always produce a better overall result than "no"—such as saving a greater number of human lives). This goes against the principles of experimental design in psychology as "the central aspects of utilitarianism and deontology—consequences and norms—are not manipulated" [16] (p. 343). Moreover, the scoring system results in mistakenly treating the utilitarian and deontological approaches to moral judgment as the two ends of a unilateral binary: the more a participant chooses "yes", the more utilitarian and less deontological they will be scored. This may be caused by confusing philosophical concepts with psychological processes. Though the utilitarian and deontological views are opposing to each other in moral philosophy, the psychological process between individuals picking each moral view when resolving a moral dilemma should not be seen as such binary opposition [16]. Instead, there are several separate independent decision-making processes [17,18] that most people would go through, but to varying degrees or in different directions, such as the sensitivity to moral norm (conceptually align with deontology) and the sensitivity to consequence (conceptually align with utilitarianism) [16].

Gawronski and colleagues [16] designed a multinominal model called the CNI model to address the above problems. Their contribution was mainly two-fold. Firstly, they designed a new set of items, including six scenarios that varied in four editions: (a) the action ("yes" answer) is moral and leads to a larger benefit; (b) the action is moral but leads to a smaller benefit; (c) the action is not moral and leads to a smaller benefit; (d) the action is not moral but leads to a larger benefit. Secondly, the new model dissociated the utilitarian–deontological inclination, yielding C (Sensitivity to Consequences) and N (Sensitivity to Norm) scores separately, plus an I score (Inaction Preference). The I score separates human beings' general tendency towards inaction when faced with moral problems [19]. This separation is necessary for psychological study on moral judgment. One typical evidence of this human tendency is that people perceived the same harm as more severe if it was caused by action rather than inaction [20].

Considering the previous example of the Nazi occupation dilemma, the answer "No" conflates the tendency towards inaction with deontological decision or sensitivity to norms. People may choose not to take an action ("No") simply because they find it difficult to decide. Similarly, the "Yes" judgment conflates participants' predisposition toward action with utilitarian decision or sensitivity to consequence. For people who answer mostly "yes" (utilitarian judgment), besides having a utilitarian inclination, it is possible that they simply choose harmful actions, regardless of their consequences. Thus, it is necessary to dissociate one's tendency of inaction from both the deontological and utilitarian tendencies, as the CNI model does.

Gawronski et al.'s [16] model offers methodological benefits. For example, with traditional moral dilemma, previous studies found that a high psychopathy level is associated with a preference for utilitarian over deontological judgment [21–23]. Further studies found the ascription for this finding might be that psychopathic participants are less likely to adverse action [23] or that they have less reduced negative emotional reactivity to harmful acts [24], which manifests as less deontological. However, the deontological tendency could not be clearly and directly addressed here as the traditional moral dilemma treats utilitarian and deontological tendencies as pure inverses of each other rather than as separated inclinations. Adding that the commonly possessed inaction preference was not taken into consideration in the traditional moral dilemma; it constrained a direct and comprehensive exploration of the fundamental rationale. Using the CNI model, scholars were able to bridge this void, revealing that that higher level of psychopathy is associated with lower levels of consequence sensitivity, norm sensitivity, and inaction preference [10,16,25].

The CNI model has been broadly adopted since its inception. In addition to the studies on moral dilemma judgment and psychopathy, researchers have employed the CNI model to examine the association between moral dilemma judgments and incidental emotion [26], foreign language [27,28], chronic stress [29], political ideology [30], and personality traits [31], just to name a few.

Despite its theoretical contribution and broad adoption, the wider application of the CNI model is hindered by one major methodological limitation: inability to generate individual-level scores [16]. Since the CNI model adopts a multinomial processing tree (MPT), which is a stochastic model that analyzes categorical data to measure cognitive processes in human behavior [32], a relatively large number of trails is required for individual measurement (for a more detailed discussion, see the "Limitation" section in [16]). In other words, parameter estimation in MPT from individual levels using 24 trails (the number of items in the CNI test) was unreliable. This drawback limits the adoption of the CNI model to between-subject designs only, and any correlational studies are not feasible. Moreover, adopting the MPT model for the CNI model estimate may not be robust enough—two studies on them showed a marginal level of statistical significance (i.e., Study 2a) and a statistically significant deviation (i.e., Study 3a) between the actual and predicted probabilities.

Gawronski and colleagues proposed a solution to enhance individual-level parameter estimation by doubling test items [10]. The drawbacks of their solution were that it was implemented in a post hoc manner and that increasing the test length may lower the quality of the responses [33]. Liu and Liao [34] made another attempt to address this limitation by changing the algorithm named CAN. Because the CAN algorithm is theoretically based on the CNI model, which uses MPT, the CAN algorithm also suffers from parameter deviation for a small number of trails. In addition, the CAN algorithm lacks a statistical index for detecting measurement errors. Liu and Liao [34] could not determine the underlying cause for a lack of correlation between the traits and estimated parameters as it remained unclear if this lack of correlation was due to measurement errors or a true lack of correlation.

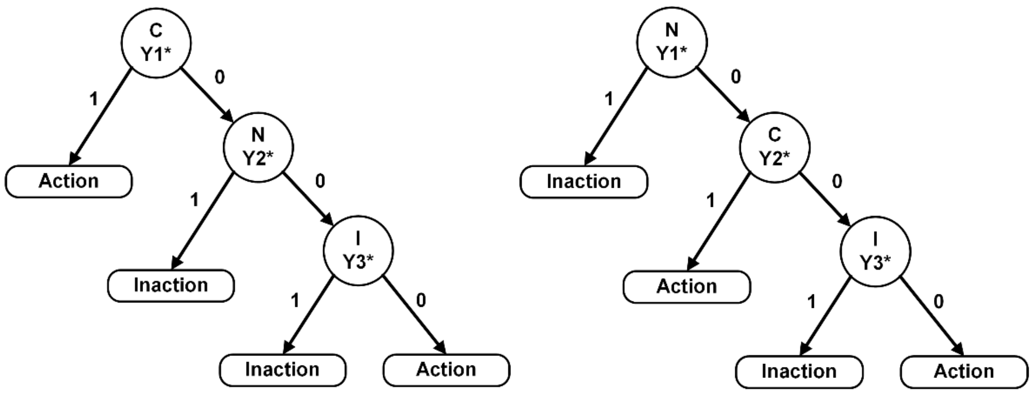
To overcome the methodological limitations of the MPT model and the CAN algorithm when analyzing the CNI moral dilemmas, the current study adopted an extended item response tree (EIRTree) model (discussed further below). As a measurement model, the EIRTree model is very flexible and has been applied in self-reported public policy decision making [35] and diagnosis of multiple strategies in educational assessments [8]. It allows for the estimation of individual scores directly, and researchers can compare different model specifications. We explored the viability of the EIRTree model with a set of published data on CNI ([10]; https://osf.io/ndf4w/ accessed on 3 July 2023) and tested the concurrent validity of scores obtained from the EIRTree model by examining its correlation with the original CNI model. The predictive validity was also tested via the association between the EIRTree-CNI model parameters and psychopathy.

2. Methods

2.1. The CNI Model

As above introduced, the CNI test includes four types of items which are the products of two dimensions: Norm (Proscriptive or Prescriptive) and Consequence (Larger or Smaller benefits compared to costs). An example of a CNI test item, the church official scenario, can be found in the beginning of Introduction. Figure 1 shows the hypothetical paths of participants' responses to each of the four types of items based on their drive to the response. This is where the implicit psychological tendencies (C, N, and I) are calculated based on explicit behavioral responses (action or inaction). By aggregating responses from multiple items, this model quantifies the extent to which the responders adopt each of three response patterns: (a) sensitive to consequence, (b) sensitive to norm, and (c) general action or inaction preference. Given different possibilities of thought processes when making a moral decision, we considered two plausible models: C→N→I and N→C→I. From a theoretical perspective, when deciding on a moral dilemma, it is equally possible that an individual first considers consequences or moral norm. However, we choose to follow Gawronski et al.'s [16] statement that "the I parameter should be set as the lowest one in the hierarchy" (p. 349) and examine only the two models rather than all six possible models involving permutations of the three variables.

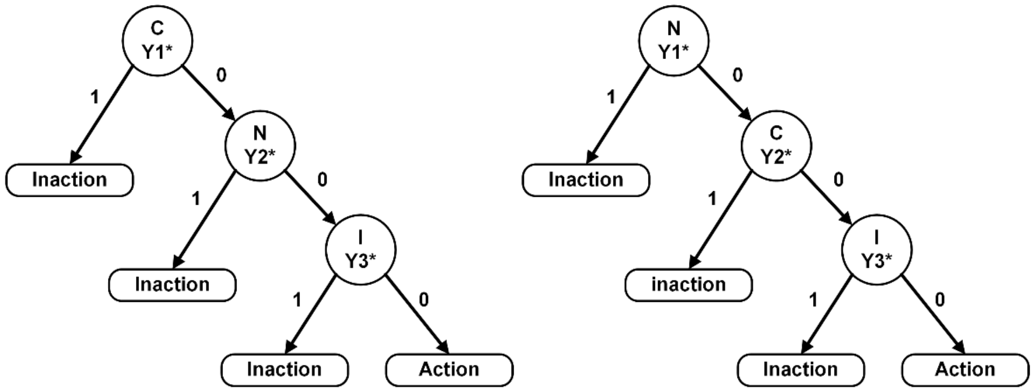
Item type 1:(progrJ) Proscriptive Norm—Benefit of action is larger than cost



Model 1 (C→N→I) Model 2 (N→C→I)

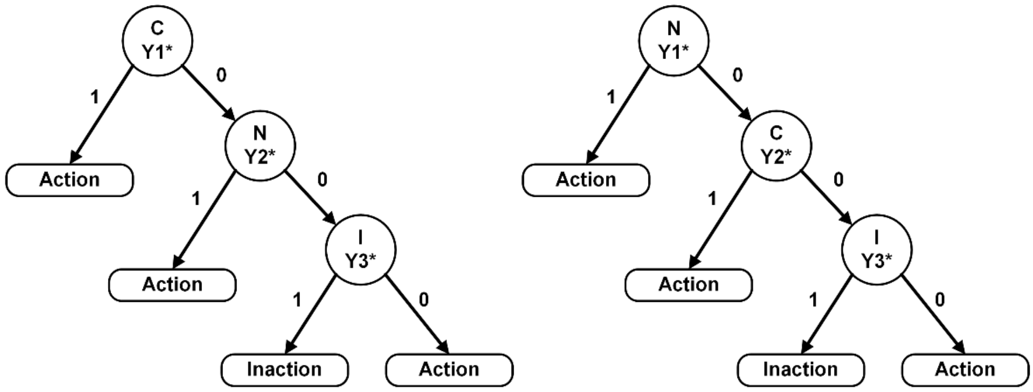
Item type 2: (prosmJ) Proscriptive Norm—Benefit of action is smaller than cost

Figure 1. *Cont.*



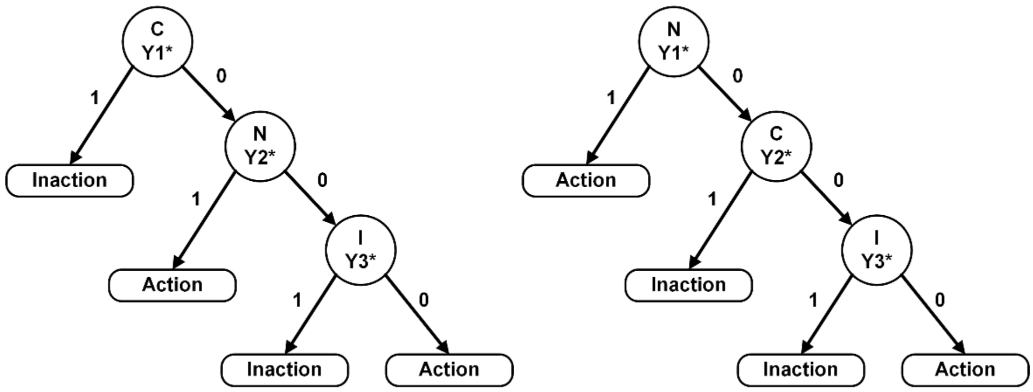
Model 1 (C→N→I) Model 2 (N→C→I)

Item type 3: (pregrJ) Prescriptive Norm—Benefit of action is larger than cost



Model 1 (C→N→I) Model 2 (N→C→I)

Item type 4: (presmJ) Prescriptive Norm—Benefit of action is smaller than cost



Model 1 (C→N→I) Model 2 (N→C→I)

Figure 1. Proposed Models.

2.2. The Proposed Approach

IRTree models are a type of IRT models that propose that the answers to polytomously scored items can be parametrized through a series of psychological processes. The categorical outcomes are modeled as a tree structure with sequentially unobserved nodes. Each node represents a subprocess with some internal outcomes that are modeled using IRT models [1]. The observed responses are the leaves, or endnotes, of the tree. Every leaf is produced through a specific path of the tree. Hence, the probability of an observed response corresponds to the probability of passing through the branches that lead to the leaf representing that response.

Take the CNI model of item type 2 (prosmJ) in Figure 1 as an example. There are three nodes, each with two internal outcomes. For the first node, denoted by Y_1^*, a respondent obtains an internal score of 1 if they choose "Inaction", or 0 if they do not make a selection but move to the second node. The CNI model assumes that the tendency of sensitivity to consequences, which is unobserved and denoted by θ_C, affects the selection of the respondent at the first node. The probability of choosing "Inaction" in the first node of item j is defined using the one-parameter logistic (1PL) IRT model as in

$$P_j\left(Y_1^* = 1 \middle| \theta_C,\, \beta_{j1}\right) = \frac{1}{1 + \exp\left(\theta_C - \beta_{j1}\right)},$$

where β_{j1} denotes item location parameter of the first node of item j. Similarly, the tendency of sensitivity to norm (θ_N score) affects the selection of the respondent at the second node, and the probability of choosing "Inaction" in the second node of item j is written by

$$P_j\left(Y_2^* = 1 \middle| \theta_N,\, \beta_{j2}\right) = \frac{1}{1 + \exp\left(\theta_N - \beta_{j2}\right)},$$

where β_{j2} denotes the item location parameter of the second node of item j. For the third node, we assume that the tendency of inaction preference (θ_I score) affects the selection of the respondent, and the probability of choosing "Inaction" is defined by

$$P_j\left(Y_3^* = 1 \middle| \theta_I,\, \beta_{j3}\right) = \frac{1}{1 + \exp\left(\theta_I - \beta_{j3}\right)},$$

where β_{j3} denotes the item location parameter of the third node of item j. Based on the above specifications, the probability of choosing "Inaction" and "Action" on item j can be calculated as

$$\begin{aligned}
P\left(X_j = \text{Inaction} \middle| \theta_C, \theta_N, \theta_I\right) &= P_j\left(Y_1^* = 1 \middle| \theta_C,\, \beta_{j1}\right) + \left[1 - P_j\left(Y_1^* = 1 \middle| \theta_C,\, \beta_{j1}\right)\right] P_j\left(Y_2^* = 1 \middle| \theta_N,\, \beta_{j2}\right) \\
&\quad + \left[1 - P_j\left(Y_1^* = 1 \middle| \theta_C,\, \beta_{j1}\right)\right]\left[1 - P_j\left(Y_2^* = 1 \middle| \theta_N,\, \beta_{j2}\right)\right] P_j\left(Y_3^* = 1 \middle| \theta_I,\, \beta_{j3}\right) \text{ and} \\
P\left(X_j = \text{Action} \middle| \theta_C, \theta_N, \theta_I\right) &= \left[1 - P_j\left(Y_1^* = 1 \middle| \theta_C,\, \beta_{j1}\right)\right]\left[1 - P_j\left(Y_2^* = 1 \middle| \theta_N,\, \beta_{j2}\right)\right]\left[1 - P_j\left(Y_3^* = 1 \middle| \theta_I,\, \beta_{j3}\right)\right]
\end{aligned}$$

respectively. For other types of items and the NCI model, the EIRTree model's specifications differ, but they can be defined in a similar way.

2.3. Data Analysis

We employed open-source data to test the proposed model ([10] https://osf.io/ndf4w/ accessed on 3 July 2023). Besides measures of moral dilemma judgment, the data also contain other variables such as participants' psychopathy level. The data consist of four item types: (i) Item type 1: proscriptive norm (progrJ; benefit of action is larger than cost), (ii) Item type 2: proscriptive norm (prosmJ; benefit of action is smaller than cost), (iii) Item type 3: prescriptive norm (pregrJ; benefit of action is larger than cost), and (iv) Item type 4: prescriptive norm (presmJ; benefit of action is smaller than cost). A total of 161 participants responded to 24 items where each item type was measured by 6 items. For each item type, we fitted the EIRTree model for the CNI and NCI.

Our data analysis was conducted in several phases. The initial phase involves transforming the data format from long to wide and recoding each data point. All item responses in the original data were assigned scores of "49" and "21" to represent action and inaction, respectively. These scores were recoded with a "1" for action and a "0" for inaction.

The parameters of the EIRTree models were estimated with the Hamiltonian Monte Carlo algorithm, which is a Markov Chain Monte Carlo (MCMC) method based on No-U-Turn sampler (NUTS) [36,37], with the rstan package (v2.30.0; [38]). The code can be downloaded from https://osf.io/a382w/ (accessed on 17 September 2023).

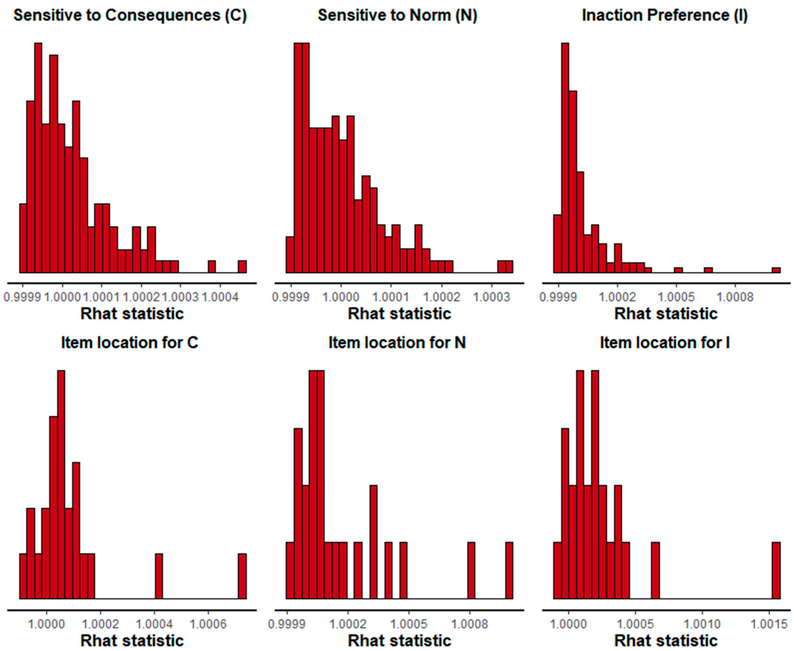
Model fits for the CNI and NCI models were examined with the chi-square (χ^2) test statistics proposed by Béguin and Glas [39]. These authors proposed the computation of two different χ^2 statistics. The first χ^2 statistic (i.e., χ_0^2) uses the observed data and the expected frequency distribution while the second χ^2 statistic (i.e., χ_{rep}^2) is derived from the frequency distribution of the replicated data and the expected frequency distribution. Using this approach, we reject the model when the posterior predictive p-value (PPP), which represents the proportion of replications where $\chi_{rep}^2 > \chi_0^2$, reaches a significantly low value.

Additionally, we conducted a comparative analysis of the two models using the Leave-One-Out cross-validation (LOO) and the Widely Applicable Information Criterion (WAIC) [40,41]. The LOO entails the sequential elimination of individual data points from the dataset, followed by assessing the model's predictive performance using the omitted data. This procedure yields significant insights into the model's capacity to generalize to new data. In contrast, WAIC is an information criterion that balances the goodness of fit and complexity of the model. Considering their predictive accuracy and complexity, LOO and WAIC provide valuable tools for selecting the most suitable model.

Further, we conducted a concurrent validity based on Pearson's bivariate correlations between scores on each new and original CNI model parameter. We followed the guidelines of [42] criteria: 0.10 (small), 0.30 (moderate), and 0.50 (large) to interpret the degree of association between these variables. The predictive validity was also examined by using Pearson bivariate correlations to test the relationship between each parameter of the new CNI model and the psychopathy level.

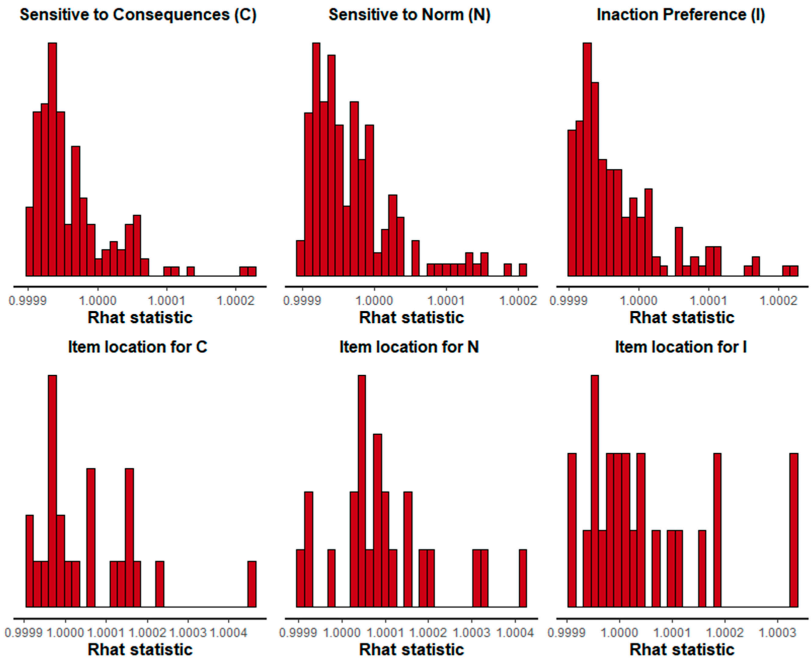
3. Results

Figure 2 displays the R-hat statistic for each model parameter. All parameters have R-hat values less than 1.01, indicating successful convergence and unbiased, consistent, and reliable MCMC estimates [43]. The chi-squared test-based posterior predictive check resulted in PPP values of 0.5278 and 0.0003 for the C→N→I and N→C→I models, respectively. Thus, the C→N→I model can fit data adequately but the N→C→I model cannot.



(**a**)

Figure 2. *Cont.*



(**b**)

Figure 2. (**a**) Rhat statistic for model 1. (**b**) Rhat statistic for model 2.

Table 1 compares the two models based on the LOO and WAIC by calculating their expected log pointwise predictive density (ELPD) for the LOO or WAIC to determine their predictive accuracy. It can be observed that the C→N→I model had a higher ELPD, suggesting that it is preferred to the N→C→I model.

Table 1. Model comparison.

	ELPD LOO		ELPD WAIC	
	Estimate	**SE**	**Estimate**	**SE**
Model 1 (CNI)	−1651.6	34.2	−1642.3	33.9
Model 2 (NCI)	−3894.1	81.0	−3585.5	75.6

Note. ELPD = expected log pointwise predictive density; SE = standard error; LOO = Leave-One-Out cross-validation; WAIC = Widely Applicable Information Criterion.

Tables 2 and 3 show the estimated location parameters for item type and item number and summarized person location estimates (i.e., tendency of sensitivity to consequences, norm, and inaction preference), respectively. The item location parameters (i.e., β_{j1}, β_{j2}, and β_{j3}) provide information on which item and node respondents are perceived as more challenging or easy. In other words, the item location parameters help determine respondents' tendency to move from one node to another or make an "Action" or "Inaction" decision. Note that items with higher location parameters may have fewer respondents making an "Action" or "Inaction" decision. Results from person location assess respondents' tendency to make decisions (θ score). Respondents with a higher θ score for sensitive to consequences, sensitive to norm, and inaction preference have higher tendencies to make moral decisions. Let us consider item 1 of progrJ (Proscriptive Norm-Benefits of action is larger than cost) as an example. In the CNI model, the mean person location for C (M = −0.26, SD = 1.58) is below the item location ($\beta_{11} = 0.37$); however, the discrepancy is relatively small (<0.5 logits). This suggests that, on average, respondents have a lower tendency to make an "action" decision on node C (sensitive to consequences) and are more likely to move to node N (sensitive to norm). At node N, the mean person location is 0.12, which is above the item location parameter ($\beta_{12} = -0.94$). This provides evidence that, on average, respondents will make an "inaction" decision at node N. However, respondents with location at node N less than the item location estimate (i.e., $\beta_{12} = -0.94$) will proceed to node I. It should be noted that the average person location at node I is 0.01, which is significantly lower than the item location, $\beta_{13} = 0.71$. This indicates that the average respondent will make an "Action" decision at node I.

Table 2. Item parameters for IRTree models.

Item Type	Item No.	CNI Model Parameters			NCI Model Parameters		
		β_{j1}	β_{j2}	β_{j3}	β_{j1}	β_{j2}	β_{j3}
progrJ: Proscriptive Norm—Benefit of action is larger than cost	1	0.37	−0.94	0.71	−0.65	1.63	−1.06
	2	2.37	−1.71	0.60	−0.04	−0.64	−0.71
	3	1.25	−1.14	1.54	−0.99	0.44	−0.67
	4	4.82	−1.19	1.83	0.54	−1.33	1.43
	5	1.76	0.45	−0.36	1.55	1.90	−0.91
	6	5.68	−0.71	−0.82	2.68	−1.45	−0.25
prosmJ: Proscriptive Norm—Benefit of action is smaller than cost	7	0.03	−0.16	2.57	−0.62	0.75	2.28
	8	2.77	−1.93	1.08	1.78	−1.64	0.92
	9	3.14	−0.86	1.54	2.07	−0.50	1.40
	10	1.42	−1.26	0.93	0.75	−0.85	0.70
	11	−0.36	−1.30	1.03	−0.73	−0.77	0.67
	12	0.90	−1.09	0.18	0.24	−0.75	0.24
pregrJ: Prescriptive Norm—Benefit of action is larger than cost	13	−0.70	0.25	−1.84	−1.25	1.24	−1.92
	14	3.73	−0.99	−1.34	2.69	−0.92	−1.47
	15	3.36	−0.24	−0.63	1.96	−0.02	−0.87
	16	−0.14	−0.84	−1.40	−0.77	−0.29	−1.47
	17	0.39	−1.22	−1.48	−0.47	−0.71	−1.70
	18	−1.34	−0.07	−2.35	−1.64	0.48	−2.35
presmJ: Prescriptive Norm—Benefit of action is smaller than cost	19	0.37	1.89	1.04	0.31	2.93	0.91
	20	1.56	−1.60	−0.96	−0.55	0.04	1.05
	21	0.80	−1.34	−1.55	−1.06	0.76	1.10
	22	5.02	−1.98	−1.84	0.12	−2.11	−1.54
	23	0.97	0.15	1.80	0.67	3.19	3.52
	24	4.52	−1.87	−2.13	−0.31	−1.82	−1.02

Note. β_{j1} = location parameter of the first node of item j; β_{j2} = location parameter of the second node of item j and β_{j3} = location parameter of the third node of item j.

Table 3. Summary of person location estimates.

	CNI Model				NCI Model		
	θ_C	θ_N	θ_I		θ_N	θ_C	θ_I
Mean	−0.26	0.12	0.01	−0.04	0.00	0.01	−0.26
Standard deviation	1.58	1.84	1.10	1.55	1.85	1.09	1.58
Minimum	−2.84	−3.74	−4.04	−3.34	−3.79	−3.87	−2.84
Maximum	4.56	2.75	2.48	3.47	3.49	2.32	4.56

Note. θ_C = person location at node C; θ_N = person location at node N and θ_I = person location at node I.

Table 4 shows the concurrent validity between the CNI scores from the original and the EIRTree models. The result shows that the estimated CNI scores from the EIRTree model have significantly strong positive correlations with the original CNI scores. This supports the concurrent validity of the EIRTree-CNI approach. Additionally, we explored the predictive validity of the CNI scores and psychopathy level (Table 4). The result suggests that psychopathy level is negatively associated with the estimated C, N, and I score from the EIRTree model, supporting predictive validity.

Table 4. Results from Pearson's bivariate correlations.

	C-Original	N-Original	I-Original	Psychopathy
C_1	0.848 ***	0.144	0.146	−0.288 ***
N_1	0.163 *	0.826 ***	0.237 **	−0.563 ***
I_1	0.147	0.233 **	0.499 ***	−0.258 **

Note. * $p < 0.05$; ** $p < 0.005$; *** $p < 0.001$; C_1, N_1 and I_1 are CNI scores from EIRTree model.

4. Discussion

The current study proposes to use the EIRTree model for the CNI test, which offers several advantages over existing methods. Firstly, it allows researchers to formally assess the fit of the model and the data, providing validity evidence about the internal structures. Secondly, it can estimate the C, N, and I scores for each individual while keeping the assessment relatively short, addressing a long-standing limitation of the CNI test. We applied the EIRTree model to a dataset that had been previously analyzed and tested for both the C→N→I and N→C→I model structures. The results indicate that the EIRTree model with the C→N→I structure fits data better than the model with the N→C→I structure. Thus, the final EIRTree-CNI model is based on the C→N→I structure, and its concurrent and predictive validity were tested and reported.

The performance of the EIRTree-CNI model over the NCI model provides several implications in the field of psychometrics and individual moral judgement that are worth considering. First, the performance of the EIRTree model applied on the CNI test suggests that the proposed model has an improved measurement accuracy. In other words, the model provides an accurate estimate of person and item parameters which help researchers to reliably distinguish between respondents with different levels of moral dilemmas. Second, insights from the CNI model can promote the development and testing of new items or calibration of existing items to scale. Additionally, the performance of the EIRTree-CNI can help researchers and moral dilemmas test developers to develop items that are more discriminating or provide better model data fit to the proposed model. It is important to note that moral decision-making processes can be challenging especially when it conflicts with respondents' personal value, social and cultural influences, ethical principles, emotional involvement, and moral development as in the stage an individual can perceive and handle moral dilemmas [44,45]. With the EIRTree-CNI model, researchers and practitioners can make informed decisions about the response processes of respondents.

While the study developed a valuable tool for future moral psychological studies, it is imperative to recognize its inherent limitations, which provide avenues for future research to explore and address. Firstly, the current study adopts a one-parameter logistic IRT model for each node because of its simplicity and smaller sample size requirement. However, researchers can use other IRT models, such as a two-parameter logistic IRT model, which considers both node threshold and discrimination parameters and may, thus, improve model–data fit. Second, the EIRTree-CNI's performance should be further examined using other datasets, along with a systematical investigation of the validity and reliability. Furthermore, a more comprehensive examination of the utility of the EIRTree model for the CNI test under varied conditions is needed. Further studies may consider using simulations to evaluate the accuracy of persons' CNI scores under different sample sizes. It is expected that larger sample sizes reduce the likelihood of bias and improve measurement precision by increasing the likelihood that estimated parameters will converge to their true values [1]. However, the determination of a sample size at which accurate estimates are obtained is beyond the scope of the current paper.

5. Concluding Remarks

The current study used an item response theory approach to score the CNI test to address its main limitation of the inability to generate individual scores without adding item numbers and employing a post hoc approach. The EIRTree-CNI model generated under such effort opens greater possibility to statistical analyses such as correlation, regression, and traditional psychometric models like classical test theory and item response theory without potential subject fatigue and compromised response quality. The subsequent test supported the model fit, concurrent validity, and predictive validity of the model.

Author Contributions: Conceptualization, Y.C. and W.M.; methodology, W.M., B.L. and Y.C.; formal analysis, W.M. and B.L.; investigation, Y.C.; data curation, W.M., B.L. and Y.C.; writing—original draft preparation, Y.C., B.L. and W.M.; writing—review and editing, Y.C., B.L., W.M. and H.H.; visualization, W.M. and B.L.; supervision, W.M. and H.H.; project administration, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data supporting the study are available at https://osf.io/ndf4w/ (accessed on 3 July 2023), and the R code is available at https://osf.io/a382w/ (accessed on 17 September 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Embretson, S.E.; Reise, S.P. *Item Response Theory for Psychologists*; L. Erlbaum Associates: Mahwah, NJ, USA, 2000.
2. Thissen, D. Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* **1982**, *47*, 175–186. [CrossRef]
3. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Danish Institute for Educational Research: Copenhagen, Denmark, 1960.
4. De Boeck, P.; Partchev, I. IRTrees: Tree-Based Item Response Models of the GLMM Family. *J. Stat. Softw. Code Snippets* **2012**, *48*, 1–28. [CrossRef]
5. Böckenholt, U.; Meiser, T. Response style analysis with threshold and multi-process IRT models: A review and tutorial. *Br. J. Math. Stat. Psychol.* **2017**, *70*, 159–181. [CrossRef] [PubMed]
6. Jeon, M.; De Boeck, P.; van der Linden, W. Modeling answer change behavior: An application of a generalized item response tree model. *J. Educ. Behav. Stat.* **2017**, *42*, 467–490. [CrossRef]
7. Ma, W. A diagnostic tree model for polytomous responses with multiple strategies. *Br. J. Math. Stat. Psychol.* **2019**, *72*, 61–82. [CrossRef]
8. Ma, W.; Guo, W. Cognitive diagnosis models for multiple strategies. *Br. J. Math. Stat. Psychol.* **2019**, *72*, 370–392. [CrossRef] [PubMed]
9. Edmonds, D. *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us about Right and Wrong*; Princeton University Press: Princeton, NJ, USA, 2013.
10. Körner, A.; Deutsch, R.; Gawronski, B. Using the CNI model to investigate individual differences in moral dilemma judgments. *Personal. Soc. Psychol. Bull.* **2020**, *46*, 1392–1407. [CrossRef]
11. McHugh, C.; McGann, M.; Igou, E.R.; Kinsella, E.L. Searching for moral dumbfounding: Identifying measurable indicators of moral dumbfounding. *Collabra Psychol.* **2017**, *3*, 23. [CrossRef]
12. Haidt, J.; Bjorklund, F.; Murphy, S. *Moral Dumbfounding: When Intuition Finds No Reason*; University of Virginia: Charlottesville, VA, USA, 2000; pp. 191–221, Unpublished manuscript.
13. Bentham, J. *A Utilitarian View, Animal Rights and Human Obligations*; Athlone: London, UK, 1789; pp. 25–26.
14. Kant, I. *Foundations of the Metaphysics of Morals*; Beck, L.W., Translator; Original work published 1785; Bobbs-Merrill: Indianapolis, IN, USA, 1959.
15. Greene, J.D.; Sommerville, R.B.; Nystrom, L.E.; Darley, J.M.; Cohen, J.D. An fMRI investigation of emotional engagement in moral judgment. *Science* **2001**, *293*, 2105–2108. [CrossRef]
16. Gawronski, B.; Armstrong, J.; Conway, P.; Friesdorf, R.; Hütter, M. Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *J. Personal. Soc. Psychol.* **2017**, *113*, 343–376. [CrossRef]
17. Conway, P.; Goldstein-Greenwood, J.; Polacek, D.; Greene, J.D. Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition* **2018**, *179*, 241–265. [CrossRef]
18. Conway, P.; Gawronski, B. Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *J. Personal. Soc. Psychol.* **2013**, *104*, 216. [CrossRef] [PubMed]
19. van den Bos, K.; Müller, P.A.; Damen, T. A Behavioral Disinhibition Hypothesis of Interventions in Moral Dilemmas. *Emot. Rev.* **2011**, *3*, 281–283. [CrossRef]
20. Spranca, M.; Minsk, E.; Baron, J. Omission and commission in judgment and choice. *J. Exp. Soc. Psychol.* **1991**, *27*, 76–105. [CrossRef]
21. Kahane, G.; Everett, J.A.; Earp, B.D.; Farias, M.; Savulescu, J. 'Utilitarian'judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition* **2015**, *134*, 193–209. [CrossRef] [PubMed]

22. Marshall, J.; Watts, A.L.; Lilienfeld, S.O. Do psychopathic individuals possess a misaligned moral compass? A meta-analytic examination of psychopathy's relations with moral judgment. *Personal. Disord. Theory Res. Treat.* **2018**, *9*, 40–50. [CrossRef]
23. Patil, I. Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion. *J. Cogn. Psychol.* **2015**, *27*, 349–366. [CrossRef]
24. Pletti, C.; Lotto, L.; Buodo, G.; Sarlo, M. It's immoral, but I'd do it! Psychopathy traits affect decision-making in sacrificial dilemmas and in everyday moral situations. *Br. J. Psychol.* **2017**, *108*, 351–368. [CrossRef]
25. Luke, D.M.; Gawronski, B. Psychopathy and moral dilemma judgments: A CNI model analysis of personal and perceived societal standards. *Soc. Cogn.* **2021**, *39*, 41–58. [CrossRef]
26. Gawronski, B.; Conway, P.; Armstrong, J.; Friesdorf, R.; Hütter, M. Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion* **2018**, *18*, 989–1008. [CrossRef]
27. Białek, M.; Paruzel-Czachura, M.; Gawronski, B. Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *J. Exp. Soc. Psychol.* **2019**, *85*, 103855. [CrossRef]
28. Białek, M.; Muda, R.; Stewart, K.; Niszczota, P.; Pieńkosz, D. Thinking in a foreign language distorts allocation of cognitive effort: Evidence from reasoning. *Cognition* **2020**, *205*, 104420. [CrossRef]
29. Zhang, L.; Kong, M.; Li, Z.; Zhao, X.; Gao, L. Chronic stress and moral decision-making: An exploration with the CNI model. *Front. Psychol.* **2018**, *9*, 1702. [CrossRef] [PubMed]
30. Luke, D.M.; Gawronski, B. Political ideology and moral dilemma judgments: An analysis using the CNI model. *Personal. Soc. Psychol. Bull.* **2021**, *47*, 1520–1531. [CrossRef]
31. Kroneisen, M.; Heck, D.W. Interindividual differences in the sensitivity for consequences, moral norms, and preferences for inaction: Relating basic personality traits to the CNI model. *Personal. Soc. Psychol. Bull.* **2020**, *46*, 1013–1026. [CrossRef]
32. Erdfelder, E.; Auer, T.S.; Hilbig, B.E.; Aßfalg, A.; Moshagen, M.; Nadarevic, L. Multinomial processing tree models: A review of the literature. *Z. Psychol./J. Psychol.* **2009**, *217*, 108–124. [CrossRef]
33. Galesic, M.; Bosnjak, M. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opin. Q.* **2009**, *73*, 349–360. [CrossRef]
34. Liu, C.; Liao, J. CAN algorithm: An individual level approach to identify consequence and norm sensitivities and overall action/inaction preferences in moral decision-making. *Front. Psychol.* **2021**, *11*, 547916. [CrossRef] [PubMed]
35. Böckenholt, U. Modeling motivated misreports to sensitive survey questions. *Psychometrika* **2014**, *79*, 515–537. [CrossRef]
36. Betancourt, M. Identifying the Optimal Integration Time in Hamiltonian Monte Carlo. *arXiv* **2016**, arXiv:1601.00225.
37. Hoffman, M.D.; Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *1*, 1539–1623.
38. Stan Development Team. RStan: The R interface to Stan. 2020. Available online: http://mc-stan.org/ (accessed on 3 July 2023).
39. Béguin, A.A.; Glas, C.A.W. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika* **2001**, *66*, 541–561. [CrossRef]
40. Luo, Y.; Jiao, H. Using the Stan program for Bayesian item response theory. *Educ. Psychol. Meas.* **2018**, *78*, 384–408. [CrossRef]
41. Vehtari, A.; Gelman, A.; Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **2015**, *27*, 1413–1432. [CrossRef]
42. Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*; Routledge: London, UK, 2013.
43. Gelman, A.; Rubin, D.B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* **1992**, *7*, 457–472. [CrossRef]
44. Haidt, J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* **2001**, *108*, 814. [CrossRef]
45. Kohlberg, L. *Stages of Moral Development as a Basis for Moral Education*; Center for Moral Education, Harvard University: Cambridge, UK, 1971; pp. 24–84.