Improved Model Exploration for the Relationship between Moral Foundations and Moral

Judgment Development Using Bayesian Model Averaging

Hyemin Han   Kelsie J. Dawson

University of Alabama

Author Note

Hyemin Han and Kelsie J. Dawson, Educational Psychology Program, University of

Alabama.

Correspondence concerning this article should be addressed to Hyemin Han,

Educational Psychology Program, University of Alabama, Tuscaloosa, AL 35487.

Contact: hyemin.han@ua.edu

The authors report no conflicting interests.

Improved Model Exploration for the Relationship between Moral Foundations and Moral

Judgment Development Using Bayesian Model Averaging

Abstract

Although some previous studies have investigated the relationship between moral

foundations and moral judgment development, the methods used have not been able to fully

explore the relationship. In the present study, we used Bayesian Model Averaging (BMA) in

order to address the limitations in traditional regression methods that have been used previously.

Results showed consistency with previous findings that binding foundations are negatively

correlated with post-conventional moral reasoning and positively correlated with maintaining

norms and personal interest schemas. In addition to previous studies, our results showed a

positive correlation for individualizing foundations and post-conventional moral reasoning.

Implications are discussed as well as a detailed explanation of the novel BMA method in order to

allow others in the field of moral education to be able to use it in their own studies.

Keywords: Bayesian Statistics; Bayesian Model Averaging; Moral Foundations; Moral

Judgment Development

Improved Model Exploration for the Relationship between Moral Foundations and Moral

Judgment Development Using Bayesian Model Averaging

Introduction

In the present study, we explored the relationship between moral foundations and moral

judgment development with the Bayesian Model Averaging (BMA) method that allows the

comparison of multiple candidate models. The theoretical frameworks of our study are based on

the Moral Foundations Theory (MFT) that proposes five different moral foundations and the

Neo-Kohlbergian theory that presents three different moral schemas. Instead of testing a specific

regression model based on a specific hypothesis, to better conduct data-driven exploration, we

employed the BMA method. We explored the BMA-identified models that predicted three

individual schemas of moral reasoning by moral foundations as independent variables and

evaluated the performance of the models. At the same time, we also introduced novel data

exploration to the community of moral education researchers for improvement of research

methodology in the field.

Moral Foundations and Moral Judgment Development

The MFT argues that the basis of moral judgment and moral decision-making are not

unidimensional but multidimensional and diverse (Graham et al., 2011). In the MFQ researchers'

original research, they proposed five different foundations for morality: harm and care (HC;

avoiding potential harms to others), fairness and reciprocity (pursuing justice in sharing; FR),

ingroup and loyalty (taking care of one's own group, family, nation; IL), authority and respect

(respecting legitimate authority and convention; AR), and purity and sanctity (avoiding

disgusting entities including actions; PS) (Graham et al., 2011; Graham, Haidt, & Nosek, 2009).

People differently value these five foundations when they consider morality and make moral

decisions. According to the MFT, people's perspectives, such as political and religious orientations, significantly influence which foundations are highly regarded. For instance, binding foundations, IL, AR, and PS, which are about binding social and community bonds, are highly valued by conservatives, while liberals more focus on individualizing foundations, HC and FR (Graham et al., 2009; Haidt & Graham, 2007). These foundations have been found to be associated with the pattern of one's moral judgment and decision-making (e.g., Vaughan, Bell Holleran, & Silver, 2019; Winterich, Zhang, & Mittal, 2012). In addition to moral judgment and decision-making, other domains in moral functioning in general, such as empathy, were also found to be associated with moral foundations. For instance, Hannikainen, Hudson, Chopik, Briley, and Derringer (2020) reported that individualizing foundations were positively correlated with empathic concern and perspective taking.

The proponents of the MFT argued that the cognitive developmental model of moral judgment, i.e., Kohlbergian and Neo-Kohlbergian theories, cannot explain the diverse nature of morality (Graham et al., 2011). According to Kohlbergian and Neo-Kohlbergian theories, moral judgment development can be explained in terms of the sophistication of post-conventional (PC) moral reasoning (Kohlberg, 1981; Rest, Narvaez, Bebeau, & Thoma, 1999a). Neo-Kohlbergians, who updated Kohlbergian theory, proposed that moral judgment development is associated with the likelihood of the utilization of three schemas, personal-interest (PI), maintaining norms (MN) and PC schemas, in making moral judgment (Rest, Narvaez, Bebeau, & Thoma, 1999b). A person with well-developed moral reasoning is likely to make judgments based on PC and universal moral principles, and less likely to rely on existing social norms or prioritize personal interests.

A widely used quantitative tool to measure one's moral judgment development is the Defining Issues Test (DIT). The DIT quantifies one's likelihood to utilize each of PC, MN, and PI schemas in moral dilemma solving. Particularly, the PC schema score, the *P*-score, has been used as an index for the development of PC reasoning (Han, Dawson, Thoma, & Glenn, 2020). Its updated version, the DIT-2, provides an additional index, the N2 score as an index for the overall sophistication of moral reasoning, which represents to what extent one prefers the PC schema against the MN and PI schema (Rest, Narvaez, Thoma, & Bebeau, 1999). Prior studies using these tools have shown that the presence of PC moral reasoning is positively associated with moral motivation and behavior (see Thoma (2002) and (2006) for review).

However, the MFT proponents criticized that such theoretical frameworks only focus on individualizing foundations, not binding foundations, in explaining the PC schema. They argued that the Kohlbergian and Neo-Kohlbergian theories are biased due to the most sophisticated moral reasoning, PC reasoning, mainly addressing individualizing values, such as justice and fairness, while valuing other important foundations relatively less (Emler, Renwick, & Malone, 1983; Graham et al., 2011). For instance, conservatives are more likely to receive lower DIT scores, particularly PC scores, compared with liberals because they highly value binding foundations but not because their moral reasoning is less developed (Graham et al., 2011). Thus, investigating the relationship between moral foundations and moral judgment development with empirical data is necessary to examine such points of concern.

Two previous studies have examined this relationship. Correlational analysis in Baril and Wright's (2012) study (Study 1) reported that the DIT PC score was negatively associated with IL and AR, while the DIT MN score was positively associated with the two foundations; the DIT PI was only positively associated with IL. In general, they found a significant association

between two binding foundations and DIT schemas. However, they could not find any direct association between moral reasoning and individualizing foundations although the (individualizing– binding) score showed significant correlation with the PC score. Glover et al. (2014) used structural equation modeling to examine the relationship between two MFQ latent variable scores, individualizing and binding foundation scores, and the DIT-2 N2 and schema scores. They reported significant association between DIT-2 scores and binding foundations. However, similar to Baril and Wright's (2012) study, individualizing foundations did not show any significant association with DIT-2 scores.

Interestingly, the aforementioned previous studies consistently reported non-significant association between individualizing foundations and moral reasoning. The PC moral reasoning is related to the consideration of universal moral principles (Kohlberg, 1981), and individualizing foundations, HC in particular, are deemed to constitute the basis of such principles by several social psychologists (Gray & Keeney, 2015; Gray & Schein, 2012; Schein & Gray, 2018). Hence, it would be generally expected that individualizing foundations, at least HC, would show meaningful relation with PC reasoning. However, the two previous studies that employed both theoretical frameworks, the MFT and moral judgment development, could not find any direct association between individualizing foundations and PC reasoning.

In addition, the methodological aspects would also warrant further investigations on the relationship between moral foundations and reasoning. Although the previous studies examined the relationship with quantitative approaches, we decided to employ a novel analysis method to address their methodological limitations. Further details about these limitations are elaborated in the following section introducing the BMA method as an alternative statistical approach.

Bayesian Model Averaging

As a way to properly explore the relationship between moral foundations and moral judgment development we employed the BMA method. The previous studies used frequentist analysis methods, which are only suitable for testing specific null hypotheses but not for model exploration. For instance, they used *P*-values to examine whether the variables of interest were significant. This is not an ideal interpretation since *p*-values can only provide information about null hypotheses, but not alternative hypotheses that researchers are mainly interested in (Wagenmakers et al., 2017). In addition, when the frequentist analysis is applied in regression, only one regression model can be tested (Raftery, 1995). Thus, even if $p < .05$ is reported, it does not necessarily mean that the tested regression model is the best model among possible alternative models consisting of possible combinations of tested independent variables. Although there are several methods to seek the best model based on the frequentist perspective, such as the stepwise regression, statisticians have warned that such methods are not ideal for the purpose because they could not properly address several critical issues such as inflation of false positives and potential multicollinearity (Hammami, Lee, Ouarda, & Lee, 2012; Harrell, 2015; Tibshirani, 1997), as well as overfitting (McNeish, 2015). Given that it is difficult to identify the best model with the traditional method, a predicted regression model is likely to be overfitted to the data used for regression. Then, the model could not well explain the reality out of the boundary of the used data.

Model selection methods based on a different perspective, Bayesian perspective, are possible ways to address the aforementioned issues. In particular, BMA is one of the most widely used Bayesian methods for model exploration (Hoeting, Madigan, Raftery, & Volinsky, 1999). The BMA process starts with assigning the prior probabilities of candidate predictors. As

a default option, the *BMA* package implemented in R employs 50% as the prior probability, P(H) (Raftery, Hoeting, Volinsky, Painter, & Yeung, 2020); this means that at the beginning, the probability to be non-zero and included in a regression model of each candidate predictor is 50%. Then, through iterative observation processes, the prior probabilities are updated with data and the posterior probabilities, P(H|D)s, are calculated with Bayes' theorem (Wagenmakers et al., 2018). P(H|D) is the probability that a specific candidate predictor is non-zero given data (Raftery, 1995). At the end of the iterative updating process, the best model can be identified by comparing the probabilities between alternative models consisting of combinations of candidate predictors (Raftery et al., 2020). As a result, a model reporting the highest probability, which includes predictors with high P(H|D)s, can be selected as the best model.

The BMA method is better than the traditional regression method for model selection in several aspects. First, it allows comparison between multiple alternative candidate models instead of testing only one model (Raftery, 1995). Related to this point, the BMA method allows us to address the issue of model uncertainty (Raftery, Madigan, & Hoeting, 1997). If we test only one single model with the traditional regression model, we cannot address model uncertainty regarding the extent to which the tested model is likely to be the true model given data (George & Clyde, 2004), as no other alternative model is tested. The BMA method provides information for the Bayesian posterior probability by comparing the multiple candidate models (George & Clyde, 2004), so we can quantitatively examine model uncertainty with this method. Second, because the Bayesian method utilizes the posterior probabilities of predictors that directly address whether the predictors are non-zero (alternative hypotheses), unlike *P*-values that only address null hypotheses, it is epistemologically better than the traditional method (Han, Park, & Thoma, 2018). Third, in general, BMA tends to penalize complex models with unnecessary

predictors, so it is robust against overfitting. Concrete examples have shown that BMA-identified models more accurately predicted dependent variables out of the boundary of data used for regression compared with the regression models predicted by ordinary linear regression (Hoeting et al., 1999). Thus, we expected that BMA would be able to address the aforementioned methodological limitations in the previous studies that examined the relationship between moral foundations and reasoning.

Present Study

In the present study, we explored the best regression models for predicting the bDIT schema scores, PC, MN, and PI scores, with moral foundations. By employing the BMA method, we examined which moral foundation variables should be included in and excluded from the best regression models. To control the effects of political and religious backgrounds, we used political and religious affiliation variables as covariates during the model selection process. Once the best models were identified by the BMA method, we examined whether the BMA-identified models were statistically better compared with the full models including all candidate predictors, which were similar to the models predicted by traditional regression methods in the prior studies. We compared Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Bayes Factor, and cross-validation performance between the two different types of models. Finally, we discussed the meanings and implications of the findings from the BMA process.

<div align="center">

Methods

</div>

Participants

Participants were college students recruited from a university located in the Southern United States of America. Participants were enrolled in courses in psychology or educational psychology and invited to sign up for an online subject pool and provided with a link to a

Qualtrics survey. When they completed the survey, a course credit was offered to them as

compensation. The design of the present study and consent form were reviewed and approved by

the Institutional Review Board.

Data collected from a total of 461 participants (14.32% male, 85.68% female) were

analyzed in the present study. Their mean age was 22.50 years (SD = 6.80 years). The

composition of the participants' political affiliations was: 44.25% Republicans, 27.77%

Democrats, 14.53% independents, 3.25% Libertarians, .22% Green Party, and 9.98% other. The

participants' religious affiliations were: 24.73% Catholics, 20.61% Evangelical Protestants,

13.02% non-Evangelical Protestants, 10.63% spiritual but not religious, 4.56% Agnostic, 3.25%

Atheistic, .65% Jewish, .43% Muslims, .22% Hindis, and 21.91% other.

Materials

Behavioral Defining Issues Test

We employed the Behavioral DIT (bDIT), which is a shortened version of the original

DIT more suitable for behavioral experiments and online studies, to examine participants' moral

judgment development (Han, Dawson, Thoma, et al., 2020). The bDIT consists of three moral

dilemmas, Heinz and the Drug, Escaped Prisoner, and Newspaper, and the overall organization is

similar to that of the original DIT. First, participants were presented with each moral dilemma.

Then, they were asked to choose their behavioral option. For instance, after presenting Heinz and

the Drug, participants were asked whether they shall steal the drug or not. After making the

behavioral choice, a total of eight items were presented that asked what the most important moral

philosophical rationale in the decision-making was. In each of the eight items, three rationale

options, one corresponding to each schema, were offered. As three dilemmas were presented, a

total of 24 items (8 per dilemma) were presented.

Then, three individual schema scores were calculated: a PC (*P*-score equivalent), MN, and PI score. Each schema score was calculated as follows:

$$Schema\ Score = \frac{\#\ of\ a\ certain\ schema\ option\ selected}{24} \times 100$$

For instance, when a participant selected PC options 12 times out of all 24 items, their PC score becomes 12 / 24 x 100 = 50. The resultant schema scores are similar to the schema scores calculated from the original DIT. For example, the bDIT PC score is conceptually equivalent to the *P*-score from the original DIT (Han, Dawson, Thoma, et al., 2020), which has been used as an indicator of sophisticated moral reasoning (Rest & Narvaez, 1994).

Previous studies testing the reliability of the bDIT showed that the internal consistency was at least acceptable (> .60) (Choi, Han, Dawson, Thoma, & Glenn, 2019; Han, Dawson, Thoma, et al., 2020). The bDIT showed good internal consistency in the present study as well (α = .77). Additionally, the internal consistency was at least acceptable in all three schemes, α = .80 in PC, α = .80 in MN, and α = .74 in PI, respectively. Moreover, the previous studies reported that the PC score calculated from the bDIT and that from the original DIT, the *P*-score, were very highly correlated with each other (Choi et al., 2019; Han, Dawson, Thoma, et al., 2020). The bDIT PC score also showed good convergent validity in terms of significant correlation with other moral psychological indicators (Han, Dawson, Choi, Choi, & Glenn, 2020), such as the moral and general growth mindset (Dweck, 2000; Han, Choi, Dawson, & Jeong, 2018), empathic concern and perspective taking (Davis, 1983), moral internalization (Aquino & Reed II, 2002), and moral disengagement (Moore, Detert, Klebe Trevino, Baker, & Mayer, 2012). Given these, the bDIT can be a reliable and valid test as a shortened version of the original in situations where the DIT is difficult to administer, such as experimental and online studies that use multiple measures.

Moral Foundations Questionnaire

We used the Moral Foundations Questionnaire (MFQ) to measure participants' scores for five moral foundations (Graham et al., 2011). This questionnaire consists of a total of 30 items that were designed to assess participants' perceived importance of each moral foundation in moral judgment. Each subscale score was calculated by averaging six items assigned to the subscale. Answers were anchored to a six-point Likert scale. The items asked whether each foundation was relevant to participants (e.g., "whether or not someone suffered emotionally"; 0: not at all relevant-5: extremely relevant) or whether they agreed with the statement about the foundation (e.g., "compassion for those who are suffering is the most crucial virtue"; 0: strongly disagree-5: strongly agree). All five subscales showed at least acceptable internal consistency: HC .70, FR .68, IL .73, AR .72, and PS .78.

Procedure

We recruited participants from psychology and educational psychology subject pool systems. They signed up for our study and received a link to our Qualtrics survey. Once they completed the consent procedure, they were randomly presented with the bDIT and MFQ. After completing the tests, participants were asked to complete a demographics survey form. Then, they were referred to the pool system again in order to receive a course credit as compensation.

Analysis

We used R to analyze the data. Before starting model exploration, we examined the descriptive statistics of the variables of interest, the bDIT schema and MFQ subscale scores. In addition, we conducted correlation analysis. Since multiple correlation coefficients were tested, the false positive discovery rate (FDR) correction was performed.

BAYESIAN MODEL AVERAGING 13

To search for the best model, we employed the BMA method implemented in an R package, *BMA*. Based on Bayesian statistics, *BMA* calculates the posterior probability of each candidate model, which indicates to which extent the model is favored by data (Raftery et al., 2020). Then, a model with the highest posterior probability was identified as the best model to be selected (Hoeting et al., 1999). We performed three BMA processes to search for the best models predicting three dependent variables, bDIT-PC, MN, and PI scores. As candidate predictors, five MFQ subscale scores were entered into the model. In addition, participants' political and religious affiliations were used as control variables and transformed into dummy variables. Once the BMA process was completed, we identified the model with the highest posterior probability as the best model.

To evaluate the identified models, we compared the BMA-identified and full models with AIC, BIC, Bayes Factors, and cross-validation performance. First, we created two linear regression models, the BMA-identified and full models, with *lm* function. Then, both AIC and BIC were calculated in order to determine which model had smaller AIC and BIC values, which is deemed better than the other (Zucchini, Claeskens, & Nguefack-Tsague, 2016).

Second, a Bayes Factor that indicates which model is relatively more probable than the other given data was employed; for instance, if two models, M_1 and M_2 are compared, a Bayes Factor BF_{12}, $P(D|M_1)/P(D|M_2)$, quantifies to which extent evidence supports M_1 against M_2 (Han, Park, et al., 2018; Wagenmakers et al., 2018). We calculated and compared Bayes Factors of the BMA-identified and full models with the *BayesFactor* package (Morey, Rouder, Jamil, Urbanek, & Ly, 2018). According to the guidelines suggested by Bayesian statisticians (Kass & Raftery, 1995), we concluded that when $BF_{BMAFull} = P(D|M_{BMA})/P(D|M_{Full}) > 3$, evidence positively supports the BMA-identified model against the full model, when $BF_{BMAFull} > 6$,

evidence strongly supports, and when $BF_{BMAFull}$ >10, evidence very strongly supports (Kass &

Raftery, 1995).

Third, we examined whether the BMA-identified model was more robust against

overfitting compared with the full models with K-fold cross-validation. K-fold cross-validation

examines whether a regression model can predict phenomena out of the boundary of the data

used for regression (Han, Lee, & Soylu, 2020). When this method is employed, $1/K$ of the whole

dataset is randomly sampled and used for validation. Then, regression is performed with the rest

of the dataset ($1-1/K$). In our study, we used 50% of the dataset for validation and the remaining

50% for regression as we employed 2-fold cross-validation. The validation dataset, which was

not used to estimate coefficients, is entered into the regression model to predict the dependent

variable. We used the mean squared error (MSE) with the validation dataset for model

evaluation. The MSE in our study was calculated as follows:

$$MSE = \frac{1}{N}\sum(\hat{y}_i - y_i)^2$$

Where N is the size of the validation dataset, $\hat{y}_i$ is the predicted value and y_i is the actual

value of the dependent variable. When two models' MSEs are compared, the smaller MSE

indicates the better model (Browne, 2000). We examined whether the MSE calculated with the

validation dataset was smaller in the BMA-identified model compared with the full model. We

repeated this procedure 10,000 times for randomization. We performed frequentist and Bayesian

t-tests to compare the models predicting PC, MN, and PI scores.

Readers interested in technical further details of statistical analyses may refer to the Open

Science Framework (OSF; https://osf.io/vtrxu/), where all data and source code files were

shared.

Results

Descriptive statistics and correlation analysis

The descriptive statistics are presented in Table 1. The result from correlation analysis was also reported in the same table.

Table 1

Descriptive statistics and correlation analysis result

	M (*SD*)	bDIT PC	bDIT MN	bDIT PI	MFQ HC	MFQ FR	MFQ IL	MFQ AR
bDIT PC	50.91 (21.44)	-						
bDIT MN	29.25 (19.33)	-.49***	-					
bDIT PI	18.99 (14.58)	-.27***	-.23***	-				
MFQ HC	4.60 (.79)	.16**	-.08	-.14*	-			
MFQ FR	4.46 (.72)	.21***	-.15*	-.11	.72***	-		
MFQ IL	3.87 (.87)	-.23**	.15*	.15*	.43***	.34***	-	
MFQ AR	4.13 (.83)	-.21***	.17**	.05	.48***	.41***	.76***	-
MFQ PS	3.99 (.96)	-.18**	.16**	.06	.45***	.39***	.65***	.73***

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. All *p*-values were corrected for FDR.

Model selection with Bayesian Model Averaging

Table 2 presents the best models predicting bDIT PC, MN, and PI scores that were selected with the BMA method. The best models reported the greatest posterior probability values. The estimated coefficients and standard errors of predictors and covariates that were included in each best model are also presented in the table.

Table 2

BMA-identified regression models

	bDIT PC	bDIT MN	bDIT PI
Variable	*B (SE)*	*B (SE)*	*B (SE)*

Intercept		35.41 (5.86)	32.25 (5.66)	23.89 (4.08)
MFQs	MFQ HC	5.47 (1.63)	-	-4.65 (.91)
	MFQ FR	7.03 (1.72)	-7.12 (1.26)	-
	MFQ IL	-6.19 (1.33)	-	4.30 (.83)
	MFQ AR	-	6.93 (1.10)	-
	MFQ PS	-4.77 (1.22)	-	-
Covariates	Independent	6.85 (2.55)	-	-
	Libertarian	-	12.12 (4.70)	-
	Non-Evangelical Protestant	10.54 (2.44)	-	-

Note. Estimated coefficients and standard errors were displayed only when respective predictors were included in the best models. Covariates other than being Independent, Libertarian, and Non-Evangelical Protestant were not selected in any best model. Reference group: Republican (political affiliation) and Catholic (religious affiliation).

Comparison with the full model

We compared the BMA-identified and the full models, which included all candidate predictors and covariates. First, AIC and BIC were compared between the two different types of models. AIC was smaller in the BMA-identified model versus the full model when bDIT PC (4,007.23 vs. 4,008.40) and PI scores were predicted (3,747.25 vs. 3,764.38); however, the full model reported the smaller AIC in predicting the bDIT MN score (3987.37 vs. 3991.13). When BIC was compared, the BMA-identified models reported the smaller BIC compared with the full models in all cases, predicting bDIT PC (4,040.29 vs. 4,095.20), MN (4,011.80 vs. 4,074.17), and PI scores (3,763.78 vs. 3,851.18).

Second, we compared Bayes Factors between two different types of models. In all cases, the BMA-identified models were very strongly supported against the full models by evidence ($BF_{BMAFull} \geq 10$). In predicting the bDIT PC, the calculated $BF_{BMAFull}$ was 1,639.85. When the

bDIT MN was predicted, $BF_{BMAFull}$ was 463.64. In the prediction of the bDIT PI, the resultant

$BF_{BMAFull}$ was 528,770.60.

Finally, we performed 2-fold cross-validation with 10,000 iterations to test whether the

BMA-identified models were more robust against overfitting compared with the full models. As

reported in Table 3, in all cases, the BMA-identified models showed smaller MSEs. Both

frequentist and Bayesian *t*-tests consistently reported that the non-zero differences were very

strongly supported by evidence.

Table 3

Results from 2-fold cross-validation

	BMA model MSE		Full model MSE					
	Mean over 10,000		Mean over 10,000					
	iterations	*SD*	iterations	*SD*	*t*	*p*	Cohen's *D*	*BF*
bDIT PC	18.76	.64	19.10	.69	-35.70	< .001	-.50	2.35E+266
bDIT MN	18.37	.68	18.78	.75	-41.33	<.001	-.58	9.75E+353
bDIT PI	14.10	.47	14.57	.51	-68.10	<.001	-.96	2.42E+903

Discussion

In the present study, we explored the best models that predicted bDIT schema scores with

five individual moral foundations using the BMA method. We controlled for the effects from

political and religious backgrounds since they are considered to be associated with one's moral

foundations and reasoning. The results showed that first, the PC schema was positively

associated with both individualizing foundations, while two binding foundations, IL and PS,

showed negative association. Second, the MN schema was positively associated with AR, but

negatively associated with FR. Third, the PI schema score showed positive correlation with IL

and negative correlation with HC. The BMA-identified models showed better performance

compared with the full models, which are employed in the frequentist analysis, in terms of

information criteria (except the AIC in the case of bDIT MN), Bayes Factors, and robustness against overfitting (cross-validation performance).

In general, findings regarding the relationship between binding foundations and moral judgment development from the present study are coherent with what has been proposed in theories in moral psychology and moral development. With the BMA method, we were able to successfully explore the specified contribution of each individual binding foundation to each schema score, which could not be well examined in the prior studies using the traditional regression method. The findings from the BMA-identified model prediction in the present study are consistent with previous studies in terms of the relationship between moral foundations and reasoning. As Baril and Wright (2012) and Glover et al. (2014) reported, we found the negative association between binding foundations, IL and PS, and the PC reasoning. The MN and PI schema scores were positively associated with AR and IL, respectively. One interesting point was that two binding foundations showed significant association with either the MN or PI score. Given that the MN schema concerns abiding by and valuing social norms and conventions while making moral judgments (Thoma, 2014), it is not surprising to see that AR, which is about following legitimate authority and respecting conventions (Haidt & Graham, 2007), was a significant predictor for the MN score. IL, on the other hand, focuses on taking care of one's own group members, such as close others (Haidt & Graham, 2007). Since the PI schema is related to prioritizing close relationships rather than larger social systems (Thoma, 2014), it is also reasonable to see that IL significantly predicted the PI schema score.

Moreover, in general, the reported associations between individualizing foundations and schema scores seem consistent with what Gray and Keeney (2015) argued about the universality of the HC foundation in moral judgment. Given that both the most and least sophisticated

schemas showed significant association with HC, HC can be considered as one of the most fundamental moral principles that has developmental implications (Schein & Gray, 2018). Also, the fact that the BMA-identified models, which is a parsimonious model with the most essential predictors (Clyde, 2000), included HC might empirically support the aforementioned point. HC negatively predicted the PI schema score perhaps due to the fact that the PI schema is mainly concerned with welfare within the boundary of one's close others, instead of potential harms to society beyond the boundary (Thoma, 2014), while HC is more about universal moral concerns (Schein & Gray, 2018). Such a trend was less obvious in the case of the MN schema, as this schema is at least related with potential harms to societal beings and conventions, which exist outside of the boundary, although the schema does not fully address universal issues (Thoma, 2014). Instead, FR negatively predicted the MN score. Given that the perspective of fairness constitutes the basis to critically deliberate upon existing norms and conventions (Emler, Tarry, & James, 2007; González, 2002), the strong perception of fairness would be negatively associated with the MN schema, which is more about maintaining existing norms instead of criticizing them (Thoma, 2014), as shown in the present study.

The aforementioned associations existed even after controlling for political and religious affiliations, which were not fully considered in the models of previous studies. Although Graham et al. (2011) argued that the Neo-Kohlbergian theory might be liberal biased and conservatives are likely to receive low DIT scores, we found pure contributions of moral foundations to the bDIT schema scores even after controlling for political and religious backgrounds. This may suggest that the variance in moral judgment development is not completely attributable to political or religious orientation but moral foundations per se explain a significant amount of the variance. It is consistent with Thoma, Narvaez, Rest, and Derryberry's (1999) argument that the

DIT provides unique information about moral judgment development "above and beyond that accounted for by … political attitudes… (p. 338)."

The present study has several methodological implications for future research as well. We employed the BMA method that allowed us to explore the best model among possible competitive candidate models and address moral uncertainty (George & Clyde, 2004; Raftery et al., 1997). The traditional regression method can only test one model against the null model, so previous studies that used the traditional method perhaps could not fully explore the relationship between variables of interest, moral foundations and reasoning. We successfully explored the best models predicting moral reasoning with individual moral foundations and control variables with the BMA method. The BMA-identified models showed better performance compared with the full models that have been widely tested with the traditional regression method. Particularly, we found improved cross-validation performance when the BMA method was implemented. Given this, the BMA methods will allow us to address the issue of overfitting that may significantly threaten the generalizability and replicability of empirical studies based on regression (Babyak, 2004). Finally, we uploaded the data and R code files, which include line-by-line comments, to our OSF project space (https://osf.io/vtrxu/) so that readers can practice the BMA method with a concrete example. By doing so, researchers in moral education who are interested in exploring the relationship between variables will be able to employ the introduced method, which is more suitable for data exploration than the traditional regression method, in their projects.

In addition, the findings from the present study might also suggest implications for moral education. Given the strong association between individualizing foundations and the post-conventional reasoning, moral educators may consider emphasizing those foundations in moral

educational activities, particularly those aiming at the development of moral judgment. Of the two individualizing foundations, HC in particular seemingly showed a consistent association with the moral judgment development as it showed positive correlation with PC and negative correlation with PI. In fact, Gray and Keeney (2015) and Gray and Schein (2012) argued that HC shall constitute the basis of moral reasoning in general. Proponents of the MFTs also acknowledged that HC is the foundation that is commonly valued across people with different political views (Graham et al., 2009). Related to this point, Greene and his colleagues, who proposed the dual-process model in moral judgment, suggested that HC can be utilized as a referencing source in solving complex moral problems that are involved in conflicts between different values and views through the deliberative process (Cushman, Young, & Greene, 2010; Greene, 2014). Given that the modern world has been highly diversified and we can see the increasing trend of conflicts between different cultures and orientations, it would be necessary to emphasize HC as the core feature in moral education (e.g., perceiving potential harms to others' wellbeing (Rest & Narvaez, 1994), caring about others' pains and concerns (Colby & Damon, 1992)) to address the aforementioned current social problem in the long term.

As a possible direction for further investigations using the bDIT, researchers may consider examining the association between automatic and behavioral aspects of moral judgment and moral foundations with the bDIT. Because the present study primarily aimed at exploring the association between moral foundations and moral judgment, we focused on the explicit outcomes of the bDIT, the schema scores. However, the bDIT has been developed to measure participants' behavioral responses to dilemmas, such as the reaction time, that can be used as proxies for automatic, implicit, and unconscious aspects of moral judgment. In fact, the first study that employed the bDIT also examined the relationship between moral judgment development, moral

competence, and behavioral responses (Han, Dawson, Thoma, et al., 2020). Thus, given that the MFT has been proposed conceptually based on the social intuitionist model (Graham et al., 2009), an approach to explain people's moral decision-making processes as automatic and intuitive processes (Haidt, 2001), examining the relationship between moral foundations and intuitive and automatic aspects of moral judgment with the bDIT would be informative for improving our knowledge about how moral judgment occurs in the reality.

There are several limitations in the present study that warrant further studies. First, although the prior research reported several preliminary findings that suggested the association between moral foundations and moral functioning in general, we only focused on moral judgment and reasoning in the present study. Given that other domains in moral functioning were out of the scope of the current study, further investigations are necessary to examine the aforementioned association. Second, we collected data from college students, so findings from the present study might not be completely generalized across different groups. Third, we used political and religious affiliation information instead of actual political and religious orientations as control variables. To address these limitations, future studies that recruit participants from diverse backgrounds and employ the multidimensional scales for political and religious orientations.

<div align="center">

References

</div>

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. doi:10.1037//0022-3514.83.6.1423

Babyak, M. A. (2004). What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosomatic Medicine*, *66*(3), 411–421. doi:10.1097/01.psy.0000127692.23278.a9

Baril, G. L., & Wright, J. C. (2012). Different types of moral cognition: Moral stages versus

moral foundations. *Personality and Individual Differences*, *53*(4), 468–473.

doi:10.1016/j.paid.2012.04.018

Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology*, *44*(1),

108–132. doi:10.1006/jmps.1999.1279

Choi, Y.-J., Han, H., Dawson, K. J., Thoma, S. J., & Glenn, A. L. (2019). Measuring moral

reasoning using moral dilemmas: evaluating reliability, validity, and differential item

functioning of the behavioural defining issues test (bDIT). *European Journal of

Developmental Psychology*, 1–10. doi:10.1080/17405629.2019.1614907

Clyde, M. (2000). Model uncertainty and health effect studies for particulate matter.

Environmetrics, *11*(6), 745–763. doi:10.1002/1099-095X(200011/12)11:6<745::AID-

ENV431>3.0.CO;2-N

Colby, A., & Damon, W. (1992). *Some do care : contemporary lives of moral commitment*. New

York, NY: Free Press.

Cushman, F., Young, L., & Greene, J. D. (2010). Our multi-system moral psychology: Towards

a consensus view. In J. D. Doris (Ed.), *The Moral Psychology Handbook* (pp. 47–71).

Oxford, UK: Oxford University Press.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a

multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–126.

doi:10.1037/0022-3514.44.1.113

Dweck, C. S. (2000). *Self-Theories: Their Role in Motivation, Personality, and Development*.

Philadephia, PA: Psychology Press.

Emler, N., Renwick, S., & Malone, B. (1983). The relationship between moral reasoning and

political orientation. *Journal of Personality and Social Psychology*, *45*(5), 1073–1080.

    doi:10.1037/0022-3514.45.5.1073

Emler, N., Tarry, H., & James, A. St. (2007). Post-conventional moral reasoning and reputation.

    Journal of Research in Personality, *41*(1), 76–89. doi:10.1016/j.jrp.2006.02.003

George, E. I., & Clyde, M. (2004). Model Uncertainty. *Statistical Science*, *19*(1), 81–94.

    doi:10.1214/088342304000000035

Glover, R. J., Natesan, P., Wang, J., Rohr, D., McAfee-Etheridge, L., Booker, D. D., … Wu, M.

    (2014). Moral rationality and intuition: An exploration of relationships between the

    Defining Issues Test and the Moral Foundations Questionnaire. *Journal of Moral

    Education*, *43*(4), 395–412. doi:10.1080/03057240.2014.953043

González, E. (2002). Defining a post-conventional corporate moral responsibility. *Journal of

    Business Ethics*, *39*, 101–108. doi:10.1023/A:1016388102599

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of

    moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046.

    doi:10.1037/a0015141

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the

    moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385.

    doi:10.1037/a0021847

Gray, K., & Keeney, J. E. (2015). Disconfirming Moral Foundations Theory on Its Own Terms.

    Social Psychological and Personality Science, *6*(8), 874–877.

    doi:10.1177/1948550615592243

Gray, K., & Schein, C. (2012). Two Minds Vs. Two Philosophies: Mind Perception Defines

    Morality and Dissolves the Debate Between Deontology and Utilitarianism. *Review of*

Philosophy and Psychology, *3*(3), 405–423. doi:10.1007/s13164-012-0112-5

Greene, J. D. (2014). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters

for Ethics. *Ethics*, *124*(4), 695–726. doi:10.1086/675875

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral

judgment. *Psychological Review*, *108*(4), 814–834.

Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral

Intuitions that Liberals may not Recognize. *Social Justice Research*, *20*(1), 98–116.

doi:10.1007/s11211-007-0034-z

Hammami, D., Lee, T. S., Ouarda, T. B. M. J., & Lee, J. (2012). Predictor selection for

downscaling GCM data with LASSO. *Journal of Geophysical Research: Atmospheres*,

117(D17), n/a-n/a. doi:10.1029/2012JD017864

Han, H., Choi, Y.-J., Dawson, K. J., & Jeong, C. (2018). Moral growth mindset is associated

with change in voluntary service engagement. *PLOS ONE*, *13*(8), e0202327.

doi:10.1371/journal.pone.0202327

Han, H., Dawson, K. J., Choi, Y. R., Choi, Y.-J., & Glenn, A. L. (2020). Development and

validation of the English version of the Moral Growth Mindset measure. *F1000Research*, *9*,

256. doi:10.12688/f1000research.23160.3

Han, H., Dawson, K. J., Thoma, S. J., & Glenn, A. L. (2020). Developmental Level of Moral

Judgment Influences Behavioral Patterns During Moral Decision-Making. *The Journal of*

Experimental Education, *88*(4), 660–675. doi:10.1080/00220973.2019.1574701

Han, H., Lee, K., & Soylu, F. (2020). Applying the Deep Learning Method for Simulating

Outcomes of Educational Interventions. *SN Computer Science*, *1*(2), 70.

doi:10.1007/s42979-020-0075-z

Han, H., Park, J., & Thoma, S. J. (2018). Why do we need to employ Bayesian statistics and how can we employ it in studies of moral education?: With practical guidelines to use JASP for educators and researchers. *Journal of Moral Education*, *47*(4), 519–537. doi:10.1080/03057240.2018.1463204

Hannikainen, I. R., Hudson, N. W., Chopik, W. J., Briley, D. A., & Derringer, J. (2020). Moral migration: Desires to become more empathic predict changes in moral foundations. *Journal of Research in Personality*, *88*, 104011. doi:10.1016/j.jrp.2020.104011

Harrell, F. E. (2015). *Regression Modeling Strategies*. Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-19425-7

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–401. doi:10.1214/ss/1009212519

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi:10.2307/2291091

Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice*. San Francisco: Harper & Row.

McNeish, D. M. (2015). Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. *Multivariate Behavioral Research*, *50*(5), 471–484. doi:10.1080/00273171.2015.1036965

Moore, C., Detert, J. R., Klebe Trevino, L., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology*, *65*(1), 1–48. doi:10.1111/j.1744-6570.2011.01237.x

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, K., & Ly, A. (2018). Package 'BayesFactor.' Retrieved from https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*,

    25, 111. doi:10.2307/271063

Raftery, A. E., Hoeting, J. A., Volinsky, C. T., Painter, I., & Yeung, K. Y. (2020). Package

    "BMA."

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian Model Averaging for Linear

    Regression Models. *Journal of the American Statistical Association*, *92*(437), 179–191.

    doi:10.1080/01621459.1997.10473615

Rest, J. R., & Narvaez, D. (1994). *Moral development in the professions: Psychology and

    applied ethics*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. J. (1999a). *Postconventional moral

    thinking: A Neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum Associates,

    Publishers.

Rest, J. R., Narvaez, D., Bebeau, M., & Thoma, S. (1999b). A Neo-Kohlbergian approach: The

    DIT and schema theory. *Educational Psychology Review*, *11*(4), 291–324.

    doi:10.1023/a:1022053215271

Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). DIT2: Devising and testing a

    revised instrument of moral judgment. *Journal of Educational Psychology*, *91*(4), 644–659.

    doi:10.1037/0022-0663.91.4.644

Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing Moral Judgment by

    Redefining Harm. *Personality and Social Psychology Review*, *22*(1), 32–70.

    doi:10.1177/1088868317698288

Thoma, S. J. (2002). An Overview of the Minnesota Approach to Research in Moral

    Development. *Journal of Moral Education*, *31*(3), 225–245.

doi:10.1080/0305724022000008098

Thoma, S. J. (2006). Research on the Defining Issues Test. In M. Killen & J. G. Smetana (Eds.), *Handbook of Moral Development* (pp. 67–91). Mahwah, NJ: Psychology Press.

Thoma, S. J. (2014). Measuring moral thinking from a neo-Kohlbergian perspective. *Theory and Research in Education*, *12*(3), 347–365. doi:10.1177/1477878514545208

Thoma, S. J., Narvaez, D., Rest, J., & Derryberry, P. (1999). Does Moral Judgment Development Reduce to Political Attitudes or Verbal Ability? Evidence Using the Defining Issues Test. *Educational Psychology Review*, *11*, 325–341. doi:10.1023/A:1022005332110

Tibshirani, R. (1997). THE LASSO METHOD FOR VARIABLE SELECTION IN THE COX MODEL. *Statistics in Medicine*, *16*(4), 385–395. doi:10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

Vaughan, T. J., Bell Holleran, L., & Silver, J. R. (2019). Applying Moral Foundations Theory to the Explanation of Capital Jurors' Sentencing Decisions. *Justice Quarterly*, *36*(7), 1176–1205. doi:10.1080/07418825.2018.1537400

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. doi:10.3758/s13423-017-1323-7

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2017). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-017-1343-3

Winterich, K. P., Zhang, Y., & Mittal, V. (2012). How political identity and charity positioning increase donations: Insights from Moral Foundations Theory. *International Journal of Research in Marketing*, *29*(4), 346–354. doi:10.1016/j.ijresmar.2012.05.002

Zucchini, W., Claeskens, G., & Nguefack-Tsague, G. (2016). Model selection. *Encyclopedia of Mathematics*. Retrieved from

http://encyclopediaofmath.org/index.php?title=Model_selection&oldid=37771