

Running head: BEHAVIORAL PATTERNS OF MORAL JUDGMENT

Developmental Level of Moral Judgment Influences Behavioral Patterns during Moral
Decision-making

Hyemin Han Kelsie J. Dawson Stephen J. Thoma
Andrea L. Glenn
University of Alabama

Author Note

Hyemin Han (hyemin.han@ua.edu), Kelsie J. Dawson
(kjdawson1@crimson.ua.edu), and Stephen J. Thoma (sthoma@ua.edu),
Educational Psychology Program, University of Alabama, Tuscaloosa, AL 35487,
USA.

Andrea L. Glenn (alglenn1@ua.edu), Center for the Prevention of Youth Behavior
Problems, University of Alabama, Tuscaloosa, AL 35487, USA.

Correspondence concerning this article should be addressed to Hyemin Han,
Educational Psychology Program, University of Alabama, Tuscaloosa, AL
35487.

Contact:

hyemin.han@ua.edu

Abstract

We developed and tested a behavioral version of the Defining Issues Test-1 revised (DIT-1r), which is a measure of the development of moral judgment. We conducted a behavioral experiment using the behavioral Defining Issues Test (bDIT) to examine the relationship between participants' moral developmental status, moral competence, and reaction time when making moral judgments. We found that when the judgments were made based on the preferred moral schema, the reaction time for moral judgments was significantly moderated by the moral developmental status. In addition, as a participant becomes more confident with moral judgment, the participant differentiates the preferred versus other schemas better particularly when the participant's abilities for moral judgment are more developed.

Keywords: moral education, moral judgment, moral reasoning, moral development, reaction time, Defining Issues Test

Developmental Level of Moral Judgment Influences Behavioral Patterns during Moral
Decision-making

Introduction

The development of moral judgment involves a developmentally ordered set of strategies used to interpret socio-moral situations (Kohlberg, 1981). In the Neo-Kohlbergian model, these strategies are defined as developmentally ordered moral schemas, spanning from adolescence into adulthood. Young adolescents attend to the personal consequences of an act as well as the consequences to individuals known to them (the personal interest schema). Throughout development, this view gives way to the maintaining norms schema, which focuses on social norms, laws, and social structures as moral criteria. In adulthood, this focus shifts to strategies that emphasize the application of shared values labeled the post-conventional schema (Rest, Narvaez, Thoma, & Bebeau, 1999). According to this model, the development of moral judgment occurs accumulatively and continuously, and a person's developmental level is explained as the likelihood of employing one of the aforementioned three schemas in moral judgment.

The most widely used measurement of the development of moral judgment based on the Neo-Kohlbergian model is the Defining Issues Test-1 revised (DIT-1r, revised 2015) (Rest, Narvaez, Bebeau, & Thoma, 1999). The full version of the DIT-1r employs six moral dilemmas (i.e., Heinz and the drug, student take-over, escaped prisoner, the doctor's dilemma, webster, and newspaper). Following each dilemma, participants rate and rank items that represent each of the three schemas (see the supplementary materials for a description of these dilemmas). The ranking data is used to identify the relative preference for each schema represented as a percentage of the total available ranks. Given the developmental ordering of the stages, the likelihood of the employment of the post-conventional schema, a P score, is often used as an overall index of development (Rest & Narvaez, 1994; Rest et al., 1999). This score has been utilized widely as an

indicator for moral judgment development in previous studies (Rest, 1990; Thoma, 2002). Previous studies have established the reliability and validity of the DIT (Rest et al., 1999; Thoma, 2002). In addition, longitudinal studies utilizing the DIT have demonstrated the overall developmental trends of moral judgment among diverse populations (Bebeau, 2002; Rest, Thoma, & Edwards, 1997).

The DIT-1r has been extensively validated and shown to have good psychometric properties as a measure of the comprehension and preference for moral schema (Thoma, 2006). Due to the psychometric qualities of the DIT-1r, it has been widely utilized in experimental studies in the field of moral education. Central to these studies has been an interest in whether educational environments can be structured to promote moral judgment development. For instance, researchers and moral educators have used this tool to evaluate moral educational programs in general (Lawrence, 1980) as well as professional ethics education programs (Bebeau, 2002), such as educational programs for engineering ethics education (Drake, Griffin, Kirkman, & Swann, 2005), computer ethics education (Staehr & Byrne, 2003), the educational contexts of college students (Maeda, Thoma, & Bebeau, 2009), and accounting ethics education (Bailey, Scott, & Thoma (2010). In addition, researchers have used the DIT to experimentally assess components of instructional programs (e.g., dilemma discussions, group projects) that are associated with changes in moral judgment (Bebeau, 2002; Schlaefli, Rest, & Thoma, 1985). Taken together, research using the DIT has reported that the level of education, particularly that of college education, and taking morality- and ethics-related courses positively contributed to the increase of P scores among diverse populations (Bebeau, 2002; Schlaefli, Rest, & Thoma, 1985).

Although the DIT has significantly contributed to empirical and quantitative research on cognitive aspects of moral development, its complicated format emphasizing rating and ranking tasks does not lend itself to studies requiring binary or multiple choice measures,

such as behavioral and functional neuroimaging studies (e.g., Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Greene, Nystrom, Engell, Darley, & Cohen, 2004) that rely on precisely measuring participants' behavioral responses. Such behavioral and neuroimaging studies require measuring participants' responses with a high temporal precision. More specifically, the examination of reaction time frequently requires test tools to detect reaction time differences on the order of 200 milliseconds (Wong, Haith, & Krakauer, 2015).

Researchers interested in examining moral decision making using behavioral indicators, such as the reaction time, have employed methods that are primarily designed to meet the demands of the experimental procedures (e.g., trolley problems). Although these relatively new methods have the advantage of identifying aspects of participants' cognitive processes that are unmeasured using indirect subjective, self-report measures (Brunken, Plass, & Leutner, 2003), they are criticized as only assessing behavioral aspects of moral decision-making, but not its developmental aspects as defined within the broader field of moral psychology (Han, 2014). Our study is an attempt to bridge this gap between research traditions by providing a measurement system that has its roots in developmental psychology but has the properties that are suitable for experimental studies requiring precise behavioral responses without a time delay.

Specifically, we developed and tested a modified version of the DIT-1r, hereafter the behavioral DIT (bDIT), while retaining the link to established developmental measures. After demonstrating a correspondence between the original DIT-1r and the bDIT, we then tested the bDIT in behavioral experiments to examine whether reaction time was associated with moral schema preference. We were particularly interested in reaction time as a behavioral indicator of participants within the context of research on moral development and moral education for several reasons. First, reaction time has been used as the most basic measurement in mental chronometry that is interested in the

measurement of cognitive speed. Such a measure has been deemed to be closely related to psychometric measures of cognitive ability, which could be difficult to directly measure (Jensen, 2006). Second, reaction time, particularly shorter reaction time, is associated with the development of various cognitive abilities (Deary, 2001; Salthouse, 2000). Hence, it can be used as an indicator of cognitive development in empirical research. Third, compared with measures strongly based on self-report, measures focusing on reaction time (e.g., the Implicit Association Test (Gattol, Sääksjärvi, & Carbon, 2011)) can be more robust against possible biases, social desirability bias in particular, which allow us to assess psychological processes more directly (Brown, Gray, & Snowden, 2009; Gannon & Rose, 2009). Because research on morality tends to be susceptible to such biases (Han, 2016), measuring reaction time will provide methodological benefits to the fields.

Several factors have been shown to influence reaction time to make moral judgments. First, reactions might be faster when judgment is made based on the most preferred schema rather than other schemas. For instance, in individuals who most prefer the post-conventional schema, reaction time to make a judgment based on post-conventional criteria should be faster than that based on the personal interest or maintaining norms criteria. Several previous psychological experiments support this assumption. When participants were asked to make a judgment that was opposite of their beliefs during syllogistic and logical reasoning tasks, the mean reaction time was significantly longer than when their judgment was consistent with their beliefs (Handley, Newstead, & Trippas, 2011; Robison & Unsworth, 2017). Likewise, people generally take a significantly longer time to make decisions while experiencing conflicts (Baron & Gürçay, 2016; Handley & Trippas, 2015).

Second, reaction times might be influenced by developmental factors. According to the Neo-Kohlbergian theory, a person with more developed abilities for moral judgment

can better understand and evaluate conflicting values and norms, and easily reject simple and biased solutions (Bebeau, 2002). Moreover, moral competence might also affect responses to moral dilemmas. Moral competence is a tendency to consistently apply a person's most preferred schema for moral judgment across different situations (Lind, 2000). This tendency is quantified in a competence score (C) measured by the Moral Competence Test (MCT), which indicates whether a person makes moral judgments consistently based on a specific schema across different situations (Lind, 2008).

Although previous studies have not found a significant association between the C score and reaction time, a neuroimaging study demonstrated a negative association between the score and activity in the anterior cingulate cortex (Prehn et al., 2008). Activity in this region has been associated with error and conflict solving, and slower reaction times (Botvinick, Cohen, & Carter, 2004; Han, Glover, & Jeong, 2014; Rushworth, Buckley, Behrens, Walton, & Bannerman, 2007).

We conducted two studies in order to develop a new behavioral measure for moral reasoning development, the bDIT, and test the association between participants' moral reasoning development and reaction time pattern. In Study 1, the bDIT was developed by extracting dilemmas and items from the original DIT-1r. Its psychometric qualities, particularly convergent validity, were assessed with online survey data. In Study 2, we conducted a behavioral experiment using the bDIT. We tested our hypotheses pertaining to the association between participants' behavioral patterns, particularly the reaction time, and developmental level of moral judgment.

While performing statistical analyses, we employed Bayesian inference in addition to classical inference. According to Peterson and Kaplan (2016), results from Bayesian inference can be more intuitively interpreted compared with those from classical inference in the studies of educational research. More specifically, the result of classical inference, a *P*-value, does not directly indicate whether an effect is statistically significant, or whether a null

or alternative hypothesis should be rejected. Instead, results from Bayesian inference can show us whether the discovered effect is meaningful and the likelihood of the acceptance of a null or alternative hypothesis. Thus, in the present study, we calculated Bayes Factors (BF) from Bayesian inference in addition to *P*-values (Han, Park, & Thoma, 2018). BF indicate how much the relative likelihood probability of a null hypothesis versus an alternative hypothesis changes after observing actual data. These values can demonstrate how strongly we should update the prior belief once data is obtained, and how strongly provided evidence supports an alternative hypothesis, presence of an actual effect, in favor of a null hypothesis.

We formulated three hypotheses based on previous studies. First, the mean reaction time when participants were making decisions coherent with their most preferred schema would be significantly shorter compared to when their decisions were not associated with the most preferred schema (a significant main effect of type of decision). Second, reaction time to make judgments based on the most preferred schema would decrease with development. Specifically, as *P* scores increase participants will spend significantly less time to make judgments based on their most preferred schema, while spending significantly longer making judgments not based on the most preferred schema (a two-way interaction: effect of type of decision by *P* score). Third, the *C* score would positively moderate the aforementioned interaction between the type of decision and *P* score (a three-way interaction effect of type of decision by *P* score by *C* score).

Study 1

In Study 1, we created the bDIT and tested its reliability and validity using online survey data.

Methods

Participants. We recruited 246 college students (200 females, $M = 18.51$ $SD = .92$; 207 Caucasians, 28 African Americans, 1 Native American or Alaska Native, 3 Asian Americans, 4 multi-ethnicities, 3 other ethnicities) at a Southern public university who

received one experimental credit as a partial fulfillment for an introductory psychology class as compensation. We posted a recruitment on the psychology subject pool webpage, and the recruited participants voluntarily signed up for our study. Then, they were provided with a link to a Qualtrics survey. All participants provided written informed consent. Of 246 participants, 87 did not pass the original DIT-1r screening process and were excluded from correlation analysis. Thus, data collected from 159 participants was used for the final correlation analysis.

Materials.

Original Defining Issues Test. To test the validity of the bDIT, we employed the original DIT-1r (Rest et al., 1999). We focused on the P score, which indicates the propensity of the employment of the post-conventional schema during moral judgment. For the present study, we used the shorter version of DIT-1r consisting of three dilemmas (i.e., Heinz and the drug, escaped prisoner, and newspaper).

Participants were asked to solve moral dilemmas such as the Heinz dilemma (see the supplementary materials for the story of escaped prisoner and newspaper), in which participants decided whether Heinz should steal an extremely expensive drug to save his wife. After making the action choice, participants were presented with a list of twelve statements containing reasons supporting the action choice. The majority of statements reflected the three moral schemas with the remaining items serving as reliability checks. For example, of the twelve statements, three were associated with the post-conventional schema (e.g., whether the law in this case is getting in the way of the most basic claim of any member of society), five with the maintaining norms schema (e.g., whether a community's laws are going to be upheld), and one with the personal interest schema (e.g., is Heinz willing to risk getting shot as a burglar or going to jail for the chance that stealing drug might help). For each statement, participants rated the extent (great, much, some, little, or no) to which the statement represented an important consideration while

making the action choice. At the end of each dilemma, participants then ranked the four most important statements (most, second most, third most, and fourth most important). These four ranked items were used to calculate the primary schema index. For example, to calculate the P score, we identified how many post-conventional (or P item) items were selected in the four ranks. If a P item was entered into the first rank it was weighted by 4. Similarly, a P item in the second rank was weighted 3, and so on for the remaining ranks for a possible story total of 10 points (i.e., 4+3+2+1). Story scores were summed across the three stories, divided by 30, and multiplied by 100 resulting in a percentage ranging from 0-100. The same process was used to compute the personal interest and maintaining norms schema scores.

Invalid DIT responses were screened out using established reliability criteria. These criteria are designed to flag individuals who completed the task without paying attention as indicated by either selecting items by the sentence complexity rather than meaning (e.g., whether the essence of living is more encompassing than the termination of dying) or by ranking (most important to fourth most important) items that were previously rated (great, much, some, little, or no) as less important. Finally, participants were flagged for failing to discriminate items as evidenced by assigning the same importance score to more than nine statements for any dilemma.

Behavioral Defining Issues Test. We created the bDIT by extracting statement items from the original DIT-1r. For each of three dilemmas, six statements were identified (two corresponding to each of the personal interest, maintaining norms, and post-conventional schemas). To determine which items to use, we examined correlations between the importance rating of each item (great to no) and the calculated score of the extent to which a specific schema was considered important across the whole test. We selected the two items showing the highest correlation for each schema from each dilemma. For this analysis, we utilized a large DIT dataset collected from 58,449 participants. Table S1

demonstrates calculated correlation coefficients and which items were selected for each dilemma. Consequently, a total of eighteen statement items were extracted.

We created a Qualtrics online survey form with the 18 extracted statement items (six per dilemma). For each dilemma, two personal interest schema options, two maintaining norms schema options, and two post-conventional schema options were selected. The Qualtrics survey presented three moral dilemmas that were identical to those in the short version of the DIT-1r in random order. Following the presentation of each dilemma story, participants were asked to make a behavioral decision similar to the original DIT-1r. Then, eight questions per dilemma were presented to participants. Each question asked which was the most important issue that informed their action choice. Three options (one personal interest, one maintaining norms, and one post-conventional) were presented in random order. Unlike the original DIT-1r that asked participants to score the importance of each statement and select the four most important statements in a dilemma, the bDIT presented participants with the aforementioned three options in a screen and asked them to select the most important option among the three. Once participants completed the whole survey, we calculated their bDIT P score by dividing the number of selected post-conventional schema statement items by 24 (the total number of questions), and multiplying the quotient by 100. The calculated P score ranged from 0 to 100.

Compared with the original DIT-1r, the bDIT was designed to have a simpler structure to enable us to collect behavioral data, reaction time in particular. Unlike the original DIT-1r that required participants to move forward and backward to answer both rating and ranking order questions while solving dilemmas, participants were only required to answer one question at a time during the bDIT session. It is possible to measure how long it takes to complete one question, reaction time, by comparing the time stamps when a question is presented versus when a participant makes a response

given the design of the bDIT. Thus, the bDIT allows us to collect behavioral data, reaction time, while collecting cognitive data, the development of moral reasoning, which is qualified in terms of a P score, simultaneously.

Procedures.

Online survey. We collected survey data, including both the original DIT-1r and bDIT data, from the participants who were recruited from the Psychology Subject Pool. Once they visited the Qualtrics survey site, they were asked to complete both DIT survey forms in random order. The median test duration (including the bDIT, original DIT-1r, and demographics survey) was 19.80 minutes (Mean = 230.33 minutes, $SD = 1531.84$ minutes). Because we could not control participants' behavior during the internet-based survey procedure, extreme outlier cases were included in the duration data; in fact, some participants spent more than 73 hours given the result from STATA's extremes command which might indicate that they did not complete the Qualtrics survey at once. Thus, we decided to use the median duration instead of the mean duration since the median value would be a better indicator to demonstrate the overall trend of the duration data containing many extreme cases.

Statistical analyses. First, we reviewed the descriptive statistics for the P scores calculated from the collected data, including mean, standard deviation, median, skewness, and kurtosis. In addition, we also tested whether the distribution of collected bDIT data was not significantly different from a normal distribution as this is an important criterion for further statistical analyses. We performed the Shapiro-Wilk test to examine the distribution of bDIT P scores.

Second, we tested the reliability and validity of the bDIT. For the reliability check, we estimated internal consistency reliability of the bDIT P score by calculating Cronbach's α . To examine whether the bDIT P score can be a good replacement of the original DIT-1r P score for behavioral experiments, we investigated the concurrent

validity of the measure. Correlation between original DIT-1r and bDIT P scores was calculated. We also examined the correlation between the P score of each moral dilemma in the original DIT-1r and the corresponding value in the bDIT. Because the calculated correlation coefficient may be diminished by the measurement error of each measure (Charles, 2005), we calculated the disattenuated correlation coefficient as well. The disattenuated correlation coefficient was calculated as follows:

$$R_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}}$$

Cronbach's α and correlation were examined with STATA 14 (STATA, 2017). In addition, we conducted the Bayesian correlation analysis with JASP 0.8.2 (Love et al., 2017; Wagenmakers, Love, et al., 2017) to test whether evidence supported the presence of significant correlation (Han et al., 2018). Due to the recent debates about P -values and frequentist methods in psychological studies (Benjamin et al., 2018), we decided to employ Bayesian methods to provide more information supporting our findings (Wagenmakers, Marsman, et al., 2017). We used the BF criteria ($2\log\text{BF} = 3$ for positive evidence, $2\log\text{BF} = 5$ for strong evidence, and $2\log\text{BF} = 10$ for very strong evidence) to determine the strength of evidence supporting presence of an effect (Han et al., 2018; Kass & Raftery, 1995).

Results

The descriptive statistics of the bDIT are presented in Table 1. The distribution of the whole group while including responses that did not pass the screening process was significantly different from a normal distribution. However, the distribution of the group only including responses that passed the screening process was not significantly different from a normal distribution given the results from Shapiro-Wilk tests. The mean of the original DIT-1r P score was 32.46, its standard deviation was 18.62, and its median was 30.00 among respondents that passed the screening process.

The calculated Cronbach's α values are presented in Table 2. These values

indicated that the bDIT possessed at least acceptable reliability (Nunnally, 1978) when cases that did not pass the screening process were excluded. The result of correlation analyses demonstrated that all bDIT and original DIT-1r P scores, including both total and dilemma-specific scores, were significantly correlated with each other except for the pair of bDIT Newspaper and original DIT-1r Heinz P scores (see Table 3). The correlation coefficient between the total bDIT and original DIT-1r P scores was .71 indicating a large effect size. The result from the Bayesian correlation analysis, $2\log BF = 107.46$, also indicated very strong evidence supporting the significant correlation. From this result, we calculated the disattenuated correlation coefficient by using the Cronbach's α value of the original DIT-1r, $\alpha = .79$, and bDIT, $\alpha = .79$. The disattenuated coefficient was calculated as follows:

$$R_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}} = \frac{.71}{\sqrt{.79}\sqrt{.79}} = .90$$

Discussion

We created the bDIT for future behavioral experiments when binary or multiple choices are required. Findings reported good psychometric qualities. The overall P scores were normally distributed. The Cronbach's α indicated at least acceptable reliability. The disattenuated correlation coefficient, .90, indicated high concurrent validity. These results suggest that the bDIT is a reliable and valid measurement for the development of moral judgment presented in a form suitable for behavioral experiments.

Study 2

In Study 2, we conducted behavioral experiments using the bDIT to examine the relationship between participants' reaction time while making a moral judgment and the developmental level of their moral judgment. In this behavioral experiment, we employed a computer program, E-Prime 2.0 (Psychology Software Tools, 2016), that allowed us to measure participants' reaction time with a higher temporal precision while

answering questions. We were able to measure participants' reaction time in the millisecond precision. We tested: whether the mean reaction time while making a decision based on the most preferred schema was shorter than while making a decision based on other schemas; second, whether the bDIT P score moderated the relationship between the decision type (whether a decision was based on the most preferred schema) and reaction time; and third, whether the aforementioned relationship among the decision type, bDIT P score, and reaction type was moderated by the MCT C score.

Methods

Participants. We recruited 108 college students (73 females, $M = 19.20$ $SD = 1.29$; 85 Caucasians, 16 African Americans, 1 Native American or Alaska Native, 2 Asian Americans, 1 Native Hawaiian or Pacific Islander, 3 other ethnicities). Similar to Study 1, they voluntarily signed up for our study after reading a recruitment posted on the psychology subject pool. All participants provided written informed consent.

Although there has not been any prior study examining the P score and reaction time, we estimated an effect size for sample size determination based on a previous study demonstrating a correlation coefficient between the P score and delinquent behavior (Thoma, Rest, & Davison, 1991). In this previous study, the calculated correlation coefficient between those two factors was $-.29$. We used G*Power to estimate a required sample size for the present study with the aforementioned correlation coefficient (Faul, Erdfelder, Lang, & Buchner, 2007). G*Power indicated that at least 72 participants were required to assure statistical power of $.80$ for correlational analysis.

In addition, we performed an additional sample size estimation to examine how many participants should be recruited to properly perform the planned regression analysis. Because we planned to perform regression analysis with a three-way interaction effect, we assumed that there were seven regressors (three main effects, three two-way interaction effects, and one three-way interaction effect) in our multiple regression

model. When a medium effect size, Cohen's $f = .29$ or Cohen's $f^2 = .15$, was entered to G*Power calculator, the calculator indicated that at least 102 participants should be recruited to assure statistical power of .80 for regression analysis.

Thus, the sample size for the present study, 108 participants, is deemed sufficient to assure the aforementioned statistical power given results from the aforementioned two sample size estimations.

Materials.

Original Defining Issues Test. We employed the original DIT-1r to test whether the bDIT can be used in lieu of the original DIT-1r as we did in Study 1.

Behavioral Defining Issues Test. We used the bDIT developed in Study 1. The calculated Cronbach's α values are presented in Table 2. Although the Cronbach's α of the Newspaper dilemmas was questionable (.63), the total score reached acceptable levels (.72). The Cronbach's α of the original DIT-1r was .68.

We performed correlation analysis to examine concurrent validity. As shown in Table 1, we excluded data collected from seven participants from the correlation analysis because they could not pass the DIT-1r screening procedure and their DIT-1r P scores were invalidated. Thus, data collected from 101 participants was used for this analysis. Correlation between the bDIT and original DIT-1r was very strong, $r = .78$, $p < .001$. The result from the Bayesian correlation analysis also demonstrated the presence of very strong evidence supporting significant correlation, $2\log\text{BF} = 79.26$. The calculated disattenuated correlation coefficient exceeded 1.00 (1.12), so it indicated good concurrent validity.

As mentioned in the methods section in Study 1, the design of the bDIT enabled us to measure participants' reaction time when they were responding to presented individual dilemma questions. Reaction time was quantified by calculating the time difference between when a specific question was presented and when a participant responded to the question with

a keyboard; each time stamp was recorded by E-Prime with the millisecond time precision.

Moral Competence Test. The MCT (Lind, 2008) measured whether participants choose a specific schema consistently and deliberately during moral judgment across different situations. It consists of two moral dilemmas: the doctor's dilemma and worker's dilemma. For each dilemma, participants were presented with twelve moral philosophical arguments explaining why the actor's behavior was morally right or wrong. Of the twelve arguments, six were in favor of the behavior (e.g., because they did not cause much damage to the company) while the remaining six were against it (e.g., because if everyone acted as the two workers did, we would be going against law and order in our society). Participants were asked to rate the extent to which they agree with each argument (-3: strongly disagree - 3: strongly agree). Each argument was designed to correspond to one of six Kohlbergian stages. Once participants completed the MCT, their C scores were calculated manually following the scoring guideline. The calculated C score ranged from 0 to 100 similar to the P score of the DIT. If a participant agreed more strongly with an argument representing a higher Kohlbergian stage consistently, she received a higher C score.

Procedures.

Behavioral experiment. Each participant was invited to a computer lab. First, the participant completed the bDIT programmed in E-Prime 2.0 (Psychology Software Tools, 2016). We designed the bDIT program to receive inputs through a keyboard. Moral dilemma stories and questions were presented in a randomized order. The reaction time for each question was calculated from the difference between the time stamp at the moment of the presentation of each question and keyboard response. After the end of the bDIT session, we asked the participant to complete the original DIT-1r and MCT.

Statistical analyses. We used STATA 14 for our analyses (STATA, 2017). We tested the reliability and validity of the bDIT by following the same procedures described in

Study 1. We sought the most preferred schema used by each participant by examining which schema was most often selected by the participant throughout the 24 questions. Then, we coded the decision type for each question by examining whether the schema selected for the question was consistent (1) or inconsistent (0) with the most preferred schema. To control for the influence of the different lengths of different stories on the reaction time, we standardized the reaction type for each dilemma story. We used the standardized reaction time following the normal distribution ($M = 0$, $SD = 1$) for our analyses.

We conducted three regression analyses to test the hypotheses. First, we performed regression analysis with the mean standardized reaction time as the dependent variable and the decision type as the independent variable. Second, we added the bDIT P score and the interaction between the bDIT P score and decision type to the first model. Third, we added the MCT C score, the interaction effect terms of the C score x decision type, C score x P score, and C score x P score x decision type. For each model, we performed Bayesian analyses with JASP 0.8.2 to examine whether the model was better than the null model and the model only with main effects (in the cases of the second and third analyses) and whether a tested effect was supported by evidence (Love et al., 2017; Wagenmakers, Love, et al., 2017).

Results

Table 1 demonstrates the descriptive statistics and result of Shapiro-Wilk test for the bDIT P score. Table 4 demonstrates results from multiple regression analyses. Our first hypothesis was not supported by the results as the main effect of decision type was not statistically significant in Model 1. The results of Bayesian analyses also showed that the regression model was not better than the null model, $2\log BF = .31$, and the effect of the decision type was also trivial, $2\log BF = .31$.

Our second hypothesis was supported by the significant interaction effect of

decision type by P score in Model 2 (see Figure 1). Our Bayesian analyses also demonstrated that the regression model with both main and interaction effects was better than the null model, $2\log BF = 15.22$, and the model only with main effects, $2\log BF = 15.22$, and there was very strong evidence supporting the presence of the significant interaction effect, $2\log BF = 14.15$. The result showed that participants with high P scores were more likely to spend a shorter time making a decision based on the most preferred schema, while spending a longer time making a decision based on the non-preferred schema; the opposite pattern was found among participants with low P scores.

Finally, our third hypothesis was also supported by the significant three-way interaction effect of decision type by P score by C score in Model 3 (see Figure 2). The findings from the Bayesian analyses also reported that Model 3 was better than the null model, $2\log BF = 11.08$, and the model only with main effects, $2\log BF = 14.85$, and the effect of the three-way interaction effect was positively supported by evidence, $2\log BF = 3.35$. As shown in Figure 2, C score influenced the relationship among mean reaction time, decision type (whether a decision was made based on the most preferred schema), and P score that was tested in Model 2 and shown in Figure 1. Among participants with high C scores, participants with high P scores were likely to spend less time to make a judgment based on the most preferred schema compared with other cases similar to the reported result from the test of Model 2 (see top of Figure 2). However, such an influence of P score on the relationship between decision type and reaction time became weaker when a participant reported a relatively low C score (see bottom of Figure 2).

Discussion

Our first hypothesis was not supported. Whether a participant made a moral decision based on the most preferred schema was not associated with the mean reaction time. Instead, our second and third hypotheses that took into account developmental psychological factors indicated by the P and C scores were supported by the findings.

The development of moral judgment and competence were significantly associated with the mean reaction time to make decisions based on the most preferred versus other schemas.

The two-way interaction effect of decision type by P score in Model 2 was significant. Sophisticated abilities for moral judgment are associated with better and easier differentiation of criteria and reasons supporting moral judgment (Bebeau, 2002), so the reaction time was perhaps significantly differentiated depending on whether a judgment was based on the most preferred schema. Opposite to our first hypothesis, whether a decision was made based on the most preferred schema alone did not significantly influence reaction time. Although previous research in the field of cognitive psychology reported that whether a decision was consistent with a participant's dominant belief and whether the decision was made without any conflict significantly influenced reaction time (Baron & Gürçay, 2016; Handley et al., 2011; Handley & Trippas, 2015; Robison & Unsworth, 2017), such a pattern was not found in our moral psychological study. This different pattern perhaps originated due to the different nature of the study. In the case of general cognitive decision-making tasks, which are unlikely to engage in philosophical value judgment, the presence of a conflict between the dominant belief set and the basis of the current decision might be the most powerful source determining reaction time (e.g., Botvinick, Cohen, & Carter, 2004; Cohen, Botvinick, & Carter, 2000; Kerns et al., 2004). However, in the case of moral decision-making, a participant is likely to be required to make value-related judgment (Han, 2014), and such a point might make her developmental status of moral reasoning significantly influence the reaction time pattern. In fact, scholars in moral development have argued that developed moral reasoning is required to make philosophically sophisticated judgment about which values should be prioritized to address moral dilemmas (Lan, Gowing, McMahon, Rieger, & King, 2008; Narvaez, 2008). Hence, the more developed moral reasoning that is

associated with abilities to better differentiate and prioritize different values might contribute to the positive association between the employment of the most preferred schema and decision-making speed as shown in Figure 1.

Moreover, moral competence (C score) significantly interacted with the aforementioned two-way interaction effect (Model 3). Lind (2008) proposed that moral competence significantly influences whether a person makes moral judgment consistently based upon a specific schema. Our finding might imply that as a participant becomes more confident with moral judgment, she differentiates the preferred versus other schemas better particularly when her abilities for moral judgment are more developed (a higher P score).

General Discussion

We aimed to develop, test, and apply the bDIT for behavioral experiments, such as neuroimaging experiments and psychological experiments intending to measure immediate and intuitive moral behavioral responses, when binary or multiple choices are required. Study 1 demonstrated that the bDIT can be a good substitute for the original DIT-1r for behavioral studies. Study 2 showed the association among participants' behavioral responses to moral dilemmas quantified in reaction time, P score measured by the bDIT, and C score measured by the MCT. The findings suggest that psychologists who are interested in both the behavioral and developmental aspects of moral judgment can use the bDIT for their experiments.

The bDIT is different from previous tools used in moral psychology in three ways. First, the traditional measures of moral development, such as the original DIT-1r, could not measure participants' behavioral patterns (e.g., the reaction time) although they can assess the participants' overall moral developmental status. Because such traditional measures have complicated structures and directions (Rest, 1990), they could not be used to detect immediate and subtle behavioral responses. Second, unlike experimental tools

used for previous behavioral and neuroimaging studies of morality, which could not evaluate participants' developmental status (e.g., Greene et al., 2004)), the bDIT can measure the likelihood of the utilization of the post-conventional schema, which is associated with developed moral judgment (Bebeau, 2002) . Our bDIT will contribute to future research examining the relationship between moral development and behavioral aspects of moral judgment by enabling researchers to measure participants' behavioral patterns with a high temporal precision. Third, in Study 1, the mean total duration of tests (including all the bDIT, original DIT-1r, and demographics survey) was shorter than twenty minutes. However, in the case of the administration of the short version of the original DIT-1r, researchers have mentioned that participants should be allotted at least twenty minutes to take only the original DIT-1r (Eynon, Nancy Thorley, & Stevens, 1997; Latif, 2001). Our bDIT requires a significantly shorter period to complete compared with the original form. Given that there has been a concern regarding the increased withdrawal rate due to such a long duration of the original DIT (Teal & Carroll, 1999), the significantly shorter duration of the bDIT will provide researchers with methodological benefits.

We also showed how participants' behavioral patterns were associated with the development of moral judgment and competence, which to our knowledge, has not been examined by previous studies. The results support assumptions about moral development and moral judgment that have not been tested at the behavioral level. First, developed moral judgment utilizing the post-conventional schema was associated with better abilities to differentiate options based on the most preferred versus other schemas. This finding might support prior research in moral development about the relationship between the development of moral reasoning and ability to better prioritize and differentiate diverse values during moral decision-making at the behavioral level (Lan et al., 2008; Narvaez, 2008). Second, moral competence positively moderated the

association between the P score and reaction time. This result supports the role of moral competence in moral judgment, enabling participants to make moral judgment based on a specific schema consistently (Lind, 2008; Prehn et al., 2008).

However, there are several limitations in the present study that warrant future work. First, in addition to Cronbach's α , additional reliability indicators, such as test-retest reliability, should be used to better examine the reliability of the bDIT. Second, although we employed the original DIT-1r to check validity, we may need to use additional moral psychological measurements (e.g., moral centrality measure (Aquino & Reed, 2002)) focusing on other aspects of moral functioning and behavior to examine how the result of the bDIT correlates with other moral psychological and behavioral indicators for additional convergent and predictive validity check. Third, we collected data only from college students; for better generalization, these experiments should be replicated with participants from diverse backgrounds. Fourth, although we examined participants' behavioral responses to the presented moral dilemmas, we could not test whether the outcome of moral judgment eventually resulted in actual moral behavior. Hence, future studies employing the bDIT should be conducted in order to address the aforementioned limitations.

References

- Aquino, K., & Reed, 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
- Bailey, C., Scott, I., & Thoma, S. (2010). Revitalizing Accounting Ethics Research in the Neo-Kohlbergian Framework: Putting the DIT into Perspective. *Behavioral Research in Accounting*, 22, 1-26.
- Baron, J., & Gürçay, B. (2016). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition*. doi:10.3758/s13421-016-0686-8

- Bebeau, M. J. (2002). The defining issues test and the four component model: contributions to professional education. *Journal of Moral Education, 31*, 271–295.
doi:10.1080/0305724022000008115
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour, 2*, 6–10. doi:10.1038/s41562-017-0189-z
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *Journal of Cognitive Neuroscience, 18*, 803–817.
doi:10.1162/jocn.2006.18.5.803
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences, 8*(12), 539–546.
- Brown, A. S., Gray, N. S., & Snowden, R. J. (2009). Implicit Measurement of Sexual Associations in Child Sex Abusers. *Sexual Abuse: A Journal of Research and Treatment, 21*(2), 166–180. doi:10.1177/1079063209332234
- Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational psychologist, 38*(1), 53-61.
- Charles, E. P. (2005). The Correction for Attenuation Due to Measurement Error: Clarifying Concepts and Creating Confidence Sets. *Psychological Methods, 10*(2), 206–226.
doi:10.1037/1082-989X.10.2.206
- Cohen, J. D., Botvinick, M., & Carter, C. S. (2000). Anterior cingulate and prefrontal cortex: who's in control? *Nature Neuroscience, 3*(5), 421–423. doi:10.1038/74783
- Deary, I. (2001). Reaction times and intelligence differences A population-based cohort study. *Intelligence, 29*(5), 389–399. doi:10.1016/S0160-2896(01)00062-9
- Drake, M. J., Griffin, P. M., Kirkman, R., & Swann, J. L. (2005). Engineering ethical curricula: Assessment and comparison of two approaches. *Journal of Engineering*

Education, 94(2), 223-231.

- Eynon, G., Nancy Thorley, H., & Stevens, K. T. (1997). Factors that influence the moral reasoning abilities of accountants: Implications for universities and the profession. *Journal of Business Ethics*, 16(12/13), 1297–1309. doi:10.1023/A:1005754201952
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. doi:10.3758/BF03193146
- Gannon, T. A., & Rose, M. R. (2009). Offense-Related Interpretative Bias in Female Child Molesters. *Sexual Abuse: A Journal of Research and Treatment*, 21(2), 194–207. doi:10.1177/1079063209332236
- Gattol, V., Sääksjärvi, M., & Carbon, C.-C. (2011). Extending the Implicit Association Test (IAT): Assessing Consumer Attitudes Based on Multi-Dimensional Implicit Associations. *PLoS ONE*, 6(1), e15849. doi:10.1371/journal.pone.0015849
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389–400. doi:10.1016/j.neuron.2004.09.027
- Han, H. (2014). Analysing theoretical frameworks of moral education through Lakatos' s philosophy of science. *Journal of Moral Education*, 43(1), 32–53. doi:10.1080/03057240.2014.893422
- Han, H. (2016). How can neuroscience contribute to moral philosophy, psychology and education based on Aristotelian virtue ethics? *International Journal of Ethics Education*, 1(2), 201–217. doi:10.1007/s40889-016-0016-9
- Han, H., Glover, G. H., & Jeong, C. (2014). Cultural influences on the neural correlate of moral decision making processes. *Behavioural Brain Research*, 259, 215–228. doi:10.1016/j.bbr.2013.11.012
- Han, H., Park, J., & Thoma, S. J. (2018). Why Do We Need to Employ Bayesian Statistics in

- Studies of Moral Education? *Journal of Moral Education*. doi:
10.1080/03057240.2018.1463204
- Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 28–43. doi:10.1037/a0021098
- Handley, S. J., & Trippas, D. (2015). Dual Processes and the Interplay between Knowledge and Structure: A New Parallel Processing Model. *Psychology of Learning and Motivation*, 62, 33–58. doi:10.1016/bs.plm.2014.09.002
- Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Amsterdam, The Netherlands: Elsevier.
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773–795. doi:10.2307/2291091
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior Cingulate conflict monitoring and adjustments in control. *Science*, 303(5660), 1023–1026. doi:10.1126/science.1089910
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice*. San Francisco: Harper & Row.
- Lan, G., Gowing, M., McMahon, S., Rieger, F., & King, N. (2008). A Study of the Relationship Between Personal Values and Moral Reasoning of Undergraduate Business Students. *Journal of Business Ethics*, 78(1–2), 121–139. doi:10.1007/s10551-006-9322-z
- Latif, D. A. (2001). The Relationship Between Ethical Reasoning and the Perception of Difficulty with Ethical Dilemmas in Pharmacy Students: Implications for Teaching Professional Ethics. *Teaching Business Ethics*, 5(1), 107–117.
doi:10.1023/A:1026502902003
- Lawrence, J. A. (1980). Moral judgment intervention studies using the Defining Issues Test.

- Journal of Moral Education*, 9(3), 178-191.
- Lind, G. (2000). The Importance of Role-Taking Opportunities for Self-Sustaining Moral Development. *Journal of Research in Education*, 10(1), 9–15.
- Lind, G. (2008). The meaning and measurement of moral judgment competence revisited: A dual-aspect model. In D. Fasko & W. Willis (Eds.), *Contemporary Philosophical and Psychological Perspectives on Moral Development and Education* (pp. 185–220). Cresskill, NJ: Hampton Press.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., & Wagenmakers, E. J. (2017). JASP (Version 0.8.2). Amsterdam, The Netherlands: Jasp project. Retrieved from <https://jasp-stats.org/>
- Maeda, Y., Thoma, S. J., & Bebeau, M. (2009). Understanding the relationship between moral judgment development and individual characteristics: The role of educational contexts. *Journal of Educational Psychology*, 101(1), 233.
- Narvaez, D. (2008). Human Flourishing and Moral Development: Cognitive and Neurobiological Perspectives of Virtue Development. In L. P. Nucci & D. Narvaez (Eds.), *Handbook of Moral and Character Education* (pp. 310–327). New York, NY: Routledge.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Peterson, S. K., & Kaplan, A. (2016). Bayesian analysis in educational psychology research: An example of gender differences in achievement goals. *Learning and Individual Differences*, 47, 129-135.
- Prehn, K., Wartenburger, I., Meriau, K., Scheibe, C., Goodenough, O. R., Villringer, A., ... Heekeren, H. R. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience*, 3(1), 33–46. doi:10.1093/Scan/Nsm037
- Psychology Software Tools. (2016). E-Prime 2.0. Pittsburgh, PA: Psychology Software

Tools.

- Rest, J. R. (1990). *DIT: Manual for the Defining Issues Test*. Minneapolis, MN: University of Minnesota Center for the Study of Ethical Development.
- 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., & Thoma, S. (1999). 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. (1999). DIT2: Devising and testing a Revised instrument of moral judgment. *Journal of Educational Psychology, 91*, 644-659.
- Rest, J. R., Thoma, S., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology, 89*(1), 5–28. doi:10.1037/0022-0663.89.1.5
- Robison, M. K., & Unsworth, N. (2017). Individual differences in working memory capacity and resistance to belief bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology, 70*(8), 1471–1484. doi:10.1080/17470218.2016.1188406
- Rushworth, M. F., Buckley, M. J., Behrens, T. E., Walton, M. E., & Bannerman, D. M. (2007). Functional organization of the medial frontal cortex. *Current Opinion in Neurobiology, 17*(2), 220–227. doi:10.1016/j.conb.2007.03.001
- Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology, 54*(1–3), 35–54. doi:10.1016/S0301-0511(00)00052-1
- Schlaefli, A., Rest, J. R., & Thoma, S. J. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the defining issues test. *Review of Educational Research, 55*(3), 319–344. doi:10.3102/00346543055003319
- Staehr, L. J., & Byrne, G. J. (2003). Using the defining issues test for evaluating computer

- ethics teaching. *IEEE Transactions on Education*, 46(2), 229-234.
- STATA. (2017). STATA/SE 14.2. College Station, TX: STATA.
- Teal, E. J., & Carroll, A. B. (1999). Moral Reasoning Skills: Are Entrepreneurs Different? *Journal of Business Ethics*, 19(3), 229–240. doi:10.1023/A:1006037510932
- 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.
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*. 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
- Wong, A. L., Haith, A. M., & Krakauer, J. W. (2015). Motor Planning. *The Neuroscientist*, 21(4), 385–398. doi:10.1177/1073858414541484

Table 1

Descriptive statistics and result of Shapiro-Wilk test for the bDIT P score

		<i>N</i>	<i>M</i>	<i>SD</i>	Median	Skewness	Kurtosis	Shapiro-Wilk <i>W</i>	<i>p</i>
Study 1	All	246	47.97	20.56	45.83	.36	2.66	.99	.02*
	DIT valid	159	49.82	22.91	50.00	.14	2.32	.99	.36
Study 2	All	108	52.51	18.69	54.17	.02	2.96	.99	.97
	DIT valid	101	52.52	18.32	54.17	.01	3.05	1.00	.98

Note. * $p < .05$.

Table 2

Reliability of the bDIT estimated by Cronbach's α in Studies 1 and 2

		α_{total}	α_{Heinz}	$\alpha_{Prisoner}$	$\alpha_{Newspaper}$
Study 1	All	.75	.74	.72	.64
	DIT valid	.79	.78	.77	.70
Study 2	All	.72	.83	.70	.63
	DIT valid	.72	.83	.70	.63

Table 3

Correlation coefficient between the bDIT and DIT-1r

	1	2	3	4	5	6	7
1. bDIT total P score							
2. bDIT Heinz P score	.63***						
3. bDIT Escaped Prisoner P score	.74***	.20*					
4. bDIT Newspaper P score	.76***	.21*	.38***				
5. DIT-1r total P score	.71***	.48***	.58***	.42***			
6. DIT-1r Heinz P score	.38***	.54***	.22**	.07	.69***		
7. DIT-1r Escaped Prisoner P score	.58***	.22**	.69***	.29***	.76***	.29***	
8. DIT-1r Newspaper P score	.61***	.32***	.40***	.59***	.75***	.25**	.39***

Note. Cases with invalid DIT-1r responses were excluded from the analysis. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4

Results of multiple regression analyses for the mean reaction time measured using the bDIT

	Mean standardized reaction time																	
	Model 1						Model 2						Model 3					
	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	ω^2	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	ω^2	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	ω^2
Decision type	.01	.03	.03	.37	.71	.00	.38	.09	.86	4.44	.000***	.08	.00	.17	.00	.00	1.00	.00
P score							.01	.00	.48	5.26	.000***	.03	.00	.00	.10	.53	.59	.00
Decision type x P score							-.01	.00	-.93	-4.57	.000***	.08	-.00	.00	-.03	-.06	.95	-.00
C score													-.02	.01	-.70	-2.43	.02*	-.00
Decision type x C score													.02	.01	1.10	2.62	.009**	.03
P score x C score													.00	.00	.79	2.39	.02*	.00
Decision type x P score x C score													-.00	.00	-1.13	-2.54	.01*	.02
<i>F</i>	<i>F</i> (1, 214) = .14						<i>F</i> (3, 212) = 9.76***						<i>F</i> (7, 208) = 5.34***					
R^2_{adj}	.00						.11						.12					

Note. Decision type: (0 = decisions not based on the most preferred schema, 1 = decisions based on the most preferred schema). * $p < .05$. **

$p < .01$. *** $p < .001$.

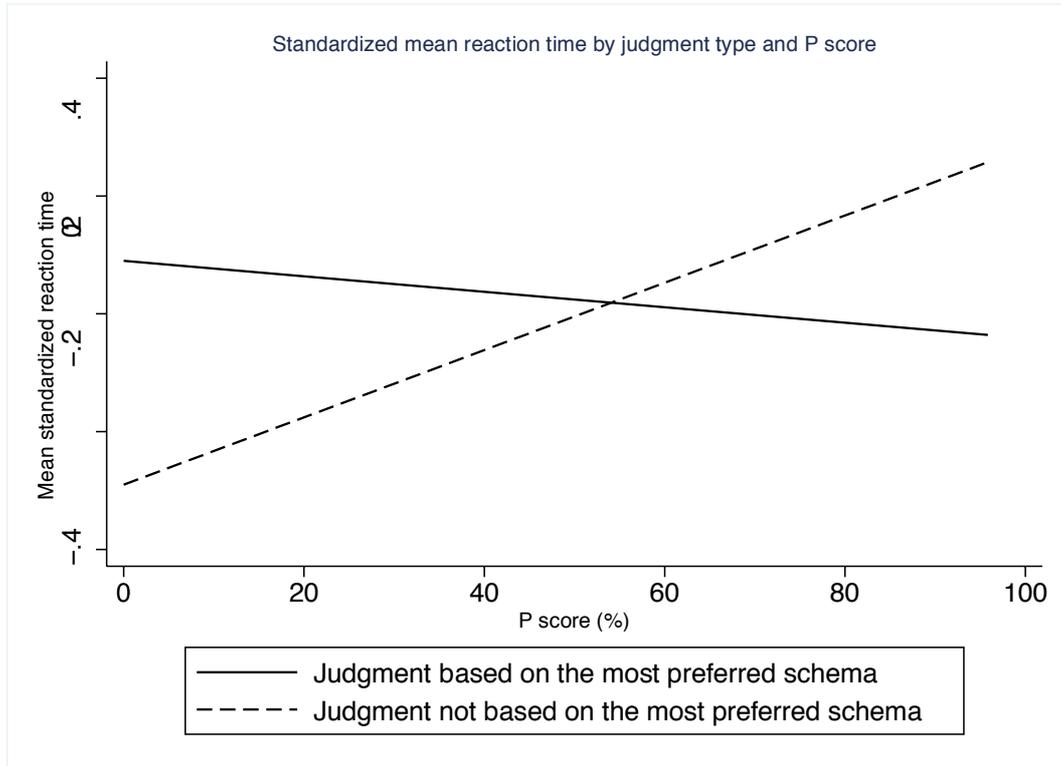


Figure 1. Standardized mean reaction time by judgment type and P score (Model 2).

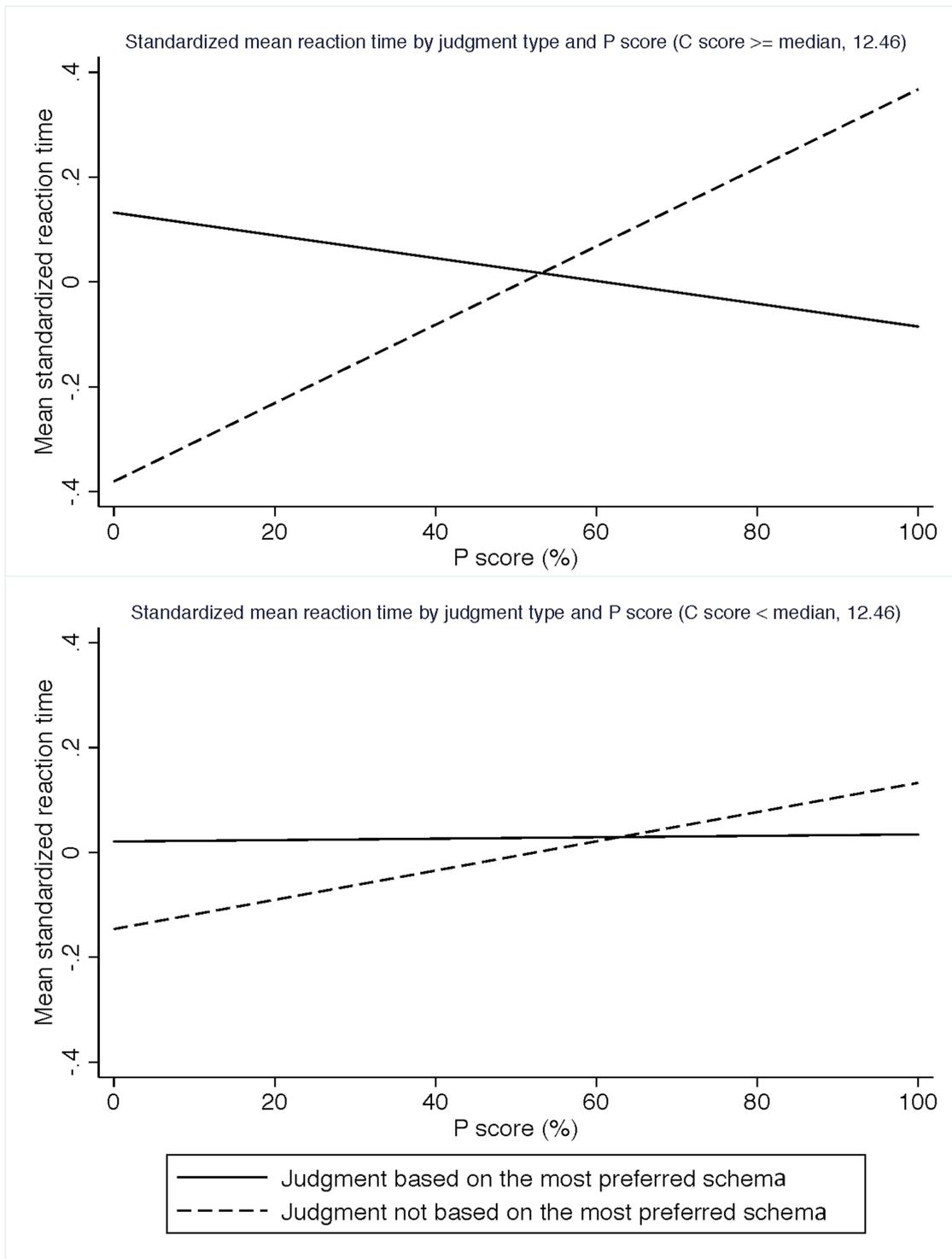


Figure 2. Standardized mean reaction time by judgment type, P score, and C score (Model 3). Top: The relationship between standardized mean reaction time, judgment type, and P score among participant with C score higher than the median. Bottom: The relationship between standardized mean reaction time, judgment type, and P score among participant with C score lower than the median.