PSYCHOMETRIKA
2012
DOI: 10.1007/S11336-012-9251-Y

THE COGNITIVE-MISER RESPONSE MODEL: TESTING FOR INTUITIVE AND DELIBERATE REASONING

ULF BÖCKENHOLT

NORTHWESTERN UNIVERSITY

In a number of psychological studies, answers to reasoning vignettes have been shown to result from both intuitive and deliberate response processes. This paper utilizes a psychometric model to separate these two response tendencies. An experimental application shows that the proposed model facilitates the analysis of dual-process item responses and the assessment of individual-difference factors, as well as conditions that favor one response tendency over another one.

Key words: local item dependencies, multiple-choice items, nominal response model, marginal maximum likelihood estimation.

> Intelligence is therefore the capacity of abstraction, which is an inhibitory process. In the intelligent moment the impulse is inhibited while it is still only partially specified, while it is still only loosely organized . . .
> (Thurstone, 1927, p. 159)

1. Introduction

Consider the following test item: "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" (Frederick, 2005). A typical immediate answer is "10 cents" because $1.10 can be divided easily into $1 and 10 cents, and 10 cents seems to be a reasonable price for a ball. However, after a moment of reflection and deliberation, a respondent may realize that the difference between $1 and 10 cents is less than $1 and give the correct answer instead. A large number of studies reported by Frederick (2005) suggest that people differ in giving a more immediate or a more deliberate answer to this question. Some respondents may rely on their intuitive reasoning, which can yield a quick and plausible judgment, while others may be more deliberate and possibly discard their immediate response impulse. Can these two sources of individual differences be disentangled based on answers to test items? To address this question, this paper presents a psychometric model, referred to as the "cognitive miser" response model, and applies it to the analysis of test items. This application shows that the proposed approach facilitates the analysis of dual-process item responses and the assessment of individual-difference factors, as well as conditions that favor one response tendency over another one.

Dual-process theories assume that there are two different modes of processing, typically referred to as System 1 and System 2 processes (Evans, 2008; Kahneman & Frederick, 2002, 2005; Stanovich, 2009). System 1 processes are characterized as unconscious, rapid, effortless and automatic, whereas System 2 processes are characterized as conscious, slow, effortful and deliberative. System 1 processes play a critical role in tasks that require immediate and parallel processing such as proprioception, depth perception, face recognition, or associate and implicit learning. In contrast, System 2 processes are the focus of our attention and are both language-

Requests for reprints should be sent to Ulf Böckenholt, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA. E-mail: u-bockenholt@kellogg.northwestern.edu

Published online: 25 January 2012

and rule-based. An important function of System 2 processing is to monitor and to potentially override System 1 processing.

Although in the context of reasoning tasks System 1 processing can yield fast, frugal, and correct conclusions, the reliance on prior beliefs and heuristics may also produce responses that are irrational. For example, with its emphasis on attributes of apparent plausibility, vividness, and salience, System 1 processing is likely to arrive at the 10 cents response for the cost of the ball in the above example. Continued thinking about the situation at hand, however, may suppress this initial incorrect response tendency. This inhibitory-control mechanism, which is a critical prerequisite for deliberate processing, was already referred to by Thurstone (1927) in his discussion on the nature of intelligence.

Much recent evidence suggests that people who are more prone to act on their System 1 response tendencies do not do so because they lack the cognitive abilities to reason correctly. For example, correlations between total SAT scores and performance on a number of vignettes that measure thinking biases have been found to be positive but modest (Frederick, 2005; Stanovich & West, 2008), accounting for less than 20% of shared variance. Correlations at the higher end of this range have been observed for vignettes that benefit from a person's knowledge about probability theory, as well as rules that support analytical thinking. The present paper builds on this work by applying a two-dimensional item-response model to test whether vignettes differ in the degree to which they elicit System 1 and System 2 processing. The model has four appealing features. First, it is likely to yield more accurate results than obtained in the literature so far because it corrects for measurement error and allows for psychometric analyses on the vignette level. Second, the model allows measuring simultaneously the separate influences of intuitive reasoning and more deliberate response processes. In particular, it shows that scoring whether incorrect responses are intuitive or deliberate provides additional information about System 1 and System 2 processes that would not be available otherwise. Third, it allows separate measurement of treatment effects with respect to the two postulated individual-difference factors, which provides further insights about the underlying response processes and validates the model structure. Fourth, the model can be used to test whether the outcomes of System 1 and System 2 processing are correlated in a population of test takers. This feature allows quantifying the extent to which inhibitory-control and cognitive abilities contribute to the quality of decision making.

The remainder of this paper is structured as follows. The next section introduces a study that was conducted to measure simultaneously the separate influences of intuitive and deliberate responses and presents some initial results. Subsequently, the modeling approach is introduced, and results obtained by fitting different versions of the model to the data are discussed. The paper concludes with a discussion of the main findings and avenues for future research.

2. Intuitive and Deliberate Responses

To explain the effects of System 1 and 2 processing in reasoning tasks, Kahneman and Frederick (2002, 2005) presented an attribute-substitution model of heuristic judgment. This model focuses on the phenomenon whereby people often answer difficult questions by substituting an answer to an easier question without necessarily being aware of the substitution. As an illustration of this heuristic, these authors refer to a survey of college students by Strack, Martin, and Schwarz (1988), who found that the correlation between the two questions "How happy are you with your life in general?" and "How many dates did you have last month?" was close to 0 when asked in the order shown but increased to 0.7 when the dating question was asked first. According to the attribute-substitution framework, the students used the dating frequency as a heuristic attribute to judge their global happiness when it was made salient by the question order. The tendency to substitute a simpler problem for a more difficult one is a general feature of many heuristic judgments. The ease with which the answer comes to mind adds to its presumed correctness ("it feels right") and provides credence to the "solution" of the attribute-substitution heuristic. Stanovich (2009) refers to people who use System 1 heuristics as a way to reduce System 2 processing as cognitive misers.

A critical assumption of this framework is that an intuitive judgment is expressed only when it is endorsed by System 2. However, this process requires cognitive resources that when reduced or depleted may lower the effectiveness of the System 2 controlling function and thus increase the probability of a System 1 response. Recently, De Neys (2006) provided support for this assumption by showing that in a secondary-task paradigm erroneous reasoning was directly caused by limitations in executive resources. More indirect evidence was provided by Bodenhausen (1990) who, building on research on time-of-day effects in human performance, found that "morning people" were more likely to give stereotypic responses in the evening and that "evening people" showed similar response tendencies when tested in the morning.

To measure the separate effects of System 1 and System 2 processing, Frederick (2005) proposed the three-item Cognitive Reflection Test (CRT). Each of these items is subject to the attribute-substitution heuristic and triggers an impulsive response that is incorrect or a more deliberate response that is correct, provided the respondent does not commit a calculational or some other reasoning error. The CRT questions (with their label in brackets) are:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? cents ["batball"]
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? minutes ["widgets"]
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? days ["lilypads"]

For the purpose of the study, these items were supplemented by two additional questions. The first question is reported by Stanovich (2009) who attributes it to Hector Levesque, and the second question was created by the author (without claiming originality):

4. Jack is looking at Anne but Anne is looking at George. Jack is married but George is not. Is a married person looking at an unmarried person? (A) Yes (B) No (C) Cannot be determined ["married"]
5. A bicyclist completes a ten-mile race track with a slow average speed of 10 miles/hour. How fast should the bicyclist cycle on the second round to reach a more impressive average speed of 20 miles/hour on both rounds? miles/hour ["speed"]

Similar to the CRT items, these questions give rise to immediate answers that need to be suppressed to obtain the correct answer. This feature is critical for the proposed modeling approach. The immediate erroneous answers to the five questions are: 10 cents, 100 minutes, 24 days, cannot be determined, and 30 miles/hour. For the first four items, the categorization of these answers as intuitive has been validated in several studies (Frederick, 2005; Stanovich, 2009). Here, we take this categorization as given and use it to test features of the proposed model. Note that with the exception of the fourth item, no response categories are offered. Thus, respondents have no external standards to verify their responses. It is also conjectured that the items can be ordered according to the degree to which they trigger an impulsive response and to the degree they are answered correctly based on a presumably deliberate response process.

As part of two online studies on a decision-making task, the five questions were presented to 462 participants. An additional 117 participants were given the five questions in a laboratory to measure their response times to each item. Most respondents were undergraduates who were paid for their participation. In addition to answering the five questions, the participants also filled out

TABLE 1.
Mean score and percentage of correct item responses by Gender and Peak Match.

Gender	Peak Match	Mean Score	Items					N
			B	W	L	M	S	
Male	Yes	2.2	42	67	65	39	4	132
Male	No	1.7	31	50	55	29	7	143
Female	Yes	1.8	35	53	58	27	2	148
Female	No	1.4	21	46	57	22	2	156

Note: The letters B, W, L, M and S stand for the respective item labels (batball, widgets, lilypads, married and speed).

the Morningness–Eveningness Questionnaire (MEQ) developed by Horne and Ostberg (1976), which was used in the Bodenhausen (1990) study. Demographic information in the form of age and gender were recorded as well. Both the online and laboratory studies were run either in the morning hours between 7 am and 10 am or in the evening hours between 7 pm and 10 pm. We selected this approach to test whether respondents exhibit more inhibitory control when asked these questions at their acrophase.

Based on their responses to the MEQ the participants were dummy coded as being at their acrophase or not. Thus, “morning people” who took the test in the morning hours and “evening people” who took the test in the evening hours fell in this category, which is referred to as peak match in the following analyses. By including gender as another factor, we obtained four groups of respondents. Table 1 contains the mean scores, as well as the percentages of correct item responses, for both gender and peak match.

The data show that, overall, male respondents seem to perform better than female respondents in answering correctly the five questions. Moreover, the data also suggest that the peak match has a substantial influence on the number of correct responses. The next section presents a model to investigate whether these differences are systematic and whether they occur at the intuitive or deliberate reasoning levels, or both.

3. A Cognitive-Miser Response (CMR) Model

The responses to each of the five items were coded into three categories. The first category captures the immediate or intuitive responses of 10 cents, 100 minutes, 24 days, cannot be determined, and 30 miles/hour for the five items, respectively. The second category contains the remaining incorrect answers that presumably are obtained because of poor execution of the deliberate reasoning process, and the third category contains the correct responses. One may question whether all responses in the second response category are obtained because of deliberate reasoning errors and apply a more fine-grained scoring instead. However, as shown below, this scoring provides already useful insights about the postulated underlying response processes.

To capture individual differences in observing the three response categories, we assume that the response process consists of two stages. In the first step, respondents arrive at the intuitive response based on System 1 processing. This response is the final answer if the respondent does not have sufficient inhibitory control to engage in systematic, effortful thought to override this response. However, if inhibitory control is available to suppress the System 1 response, deliberate processing can revise the intuitive response and arrive at the correct answer, provided no algorithmic or other calculational errors are committed.

This simple two-stage process captures some of the key characteristics of the interplay between System 1 and System 2 processing: The quick initial answer which may be suppressed by inhibitory control and the subsequent revision based on more deliberate processing. However, in the same way as the impulsive response process may not lead to the correct answer if the attribute-substitution heuristic is inappropriately applied, the deliberate response process may fail as well for faulty execution or application of logical or mathematical rules. Importantly, both processing systems can commit different errors, which form an observable basis for distinguishing them. In the considered application, each of the five items allows distinguishing between these processes because the intuitive responses are assumed to be known a priori. In general, however, this may not be the case. Seemingly intuitive errors may be the result of System 2 processes and seemingly deliberate errors may result from System 1 processes. In this case, additional process measures may be needed to distinguish the responses of both systems.

The tree diagram in Figure 1 depicts the hypothesized response process. At the first level, we postulate an inhibitory-control mechanism that determines whether an (incorrect) intuitive response (labeled as "E1") is given or whether a deliberate response process is activated. The output of the deliberate response process is either incorrect (response category "E2") or correct (response category "R"). Substantial individual differences can be expected at either stage. People differ in their ability to inhibit actions (Logan, Schachar, & Tannock, 1997) as well in their ability to reason. These individual differences in inhibitory control and deliberate processing are represented by the two person parameters $\theta^{(C)}$ and $\theta^{(D|C)}$ in Figure 1. As a first step, we assume that the two postulated response tendencies can be captured by the Rasch model (Rasch, 1960). Thus, the probability that person i gives an "E1" response can be written as

$$\Pr(x_{ij} = \text{“E1”}) = 1 - \Psi\bigl(\theta_i^{(C)} + \gamma_j^{(C)}\bigr) = 1 - \Pr{}_j\bigl(\theta_i^{(C)}\bigr), \tag{1}$$

where $\gamma_j^{(C)}$ is a location parameter capturing the difficulty for respondents to inhibit their impulsive response tendency for item j and Ψ is the cumulative logistic distribution function.

To represent the probability of an "R" response, we assume further that the test taker does not yield to the impulse of the immediate response and is able to move on to a more deliberate reasoning process. In this case, the probability that person i gives the answer "R" can be written as

$$\Pr(x_{ij} = \text{“R”}) = \Pr{}_j\bigl(\theta_i^{(C)}\bigr)\Pr{}_j\bigl(\theta_i^{(D|C)}\bigr) = \Psi\bigl(\theta_i^{(C)} + \gamma_j^{(C)}\bigr)\Psi\bigl(\theta_i^{(D|C)} + \gamma_j^{(D|C)}\bigr), \tag{2}$$

where $\gamma_j^{(D|C)}$ is the item location capturing the deliberate response difficulty for item j after the inhibitory-control stage. Consequently, the probability of an "E2" response is given by the product of the probabilities of not giving in to the immediate response and of not arriving at the correct response after deliberation:

$$\Pr(x_{ij} = \text{“E2”}) = \Pr{}_j\bigl(\theta_i^{(C)}\bigr)\bigl(1 - \Pr{}_j\bigl(\theta_i^{(D|C)}\bigr)\bigr). \tag{3}$$

The two person parameters, $\theta_i^{(C)}$ and $\theta_i^{(D|C)}$, are specified to follow a bivariate normal distribution with mean vector $\mathbf{0}$. The items, as well as the individual-difference parameters, can be related to explanatory covariates (De Boeck & Wilson, 2004). Specifically, $\theta_i^{(C)}$ and $\theta_i^{(D|C)}$ can be expressed as a function of covariates with $\theta_i^{(C)} = \sum_f^{(C)} z_{if}\beta_f^{(C)} + \epsilon_i^{(C)}$ and similar notation for $\theta_i^{(D|C)}$. Thus, random effects that are not captured by the available covariates are represented by $\epsilon_i^{(C)}$ and $\epsilon_i^{(D|C)}$, which are specified to be normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$.

The representation given by (1) to (3) is referred to as the cognitive-miser response (CMR) model. The CMR model is related to the nominal categories model (Bock, 1972; Thissen & Steinberg, 1984) in that it captures differences among three nominal response categories but varies in the important respects that the CMR model is two-dimensional and that the selection of the response categories is assumed to be sequential and not simultaneous (Suh & Bolt, 2010).

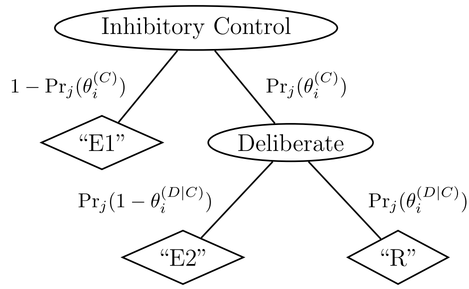

FIGURE 1.
Tree diagram of cognitive-miser model.

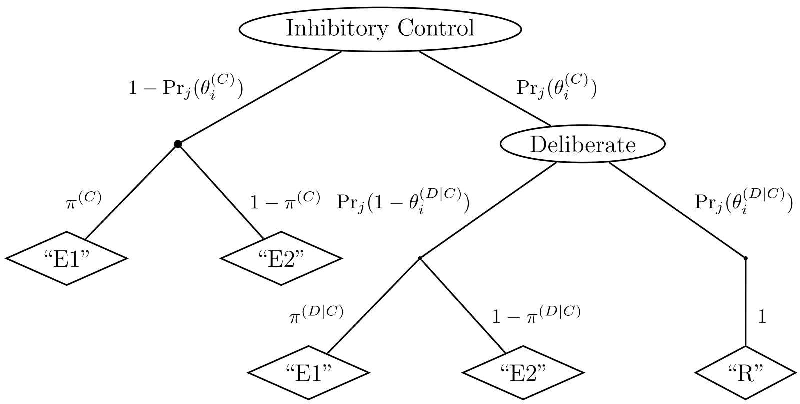

FIGURE 2.
Tree diagram of extended cognitive-miser Model A.

The model can also be viewed as a two-dimensional version of a steps model (Verhelst, Glass, & de Vries, 1997; Tutz, 1997).

In the reported application, we compare the CMR model to two extended versions depicted in Figures 2 and 3 to test whether the response-type classification agrees with the postulated response mechanisms. Both models allow for the possibility that participants may arrive at E1 and E2 responses based on intuitive reasoning. The models also assume that if a respondent can inhibit reporting the E1 responses and continues with a more deliberate reasoning process, it is still possible that this person will not arrive at the correct answer. However, the models differ in the underlying response mechanism for an E1 or E2 answer. Under Model A, respondents can give E1 or E2 responses with fixed probability at either the intuitive or deliberate reasoning stages. Under Model B, the probability of selecting one of the two incorrect response types is assumed to vary across test takers.

As can be seen from Figure 2, Model A is a straightforward extension of the CMR model, with two added parameters, $\pi^{(C)}$ and $\pi^{(D|C)}$, that capture the probability of an E1 response under intuitive and deliberate reasoning, respectively. The parameters can be interpreted as misclassification probabilities in the sense that an E2 response should have been coded as an E1 response at the intuitive stage, or that an E1 response should have been coded as an E2 response at the deliberate stage. Model A simplifies to the CMR model when $\pi^{(C)} = 1$ and $\pi^{(D|C)} = 0$.

Model B assumes that a respondent may arrive at an E2-type response both via intuitive and deliberate reasoning. Because there are three possible outcomes at the first processing stage, we

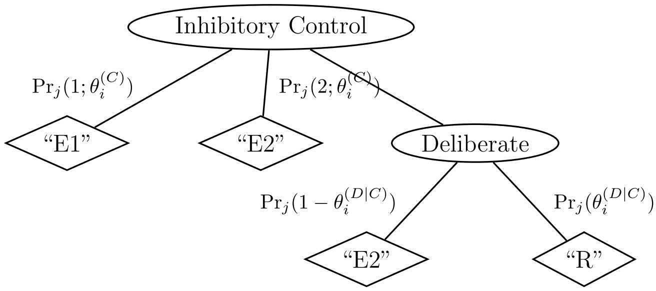

FIGURE 3.
Tree diagram of extended cognitive-miser Model B.

use the nominal categories model (Bock, 1972) to represent the probabilities of these outcomes. Specifically, the probability of an E1 response can be written as

$$\Pr(x_{ij} = \text{“E1”}) = \frac{\exp(\theta_i^{(C)} + \gamma_{j(\text{E1})}^{(C)})}{\exp(\theta_i^{(C)} + \gamma_{j(\text{E1})}^{(C)}) + \exp(\theta_i^{(C)} + \gamma_{j(\text{E2})}^{(C)}) + 1} = \Pr_j\big(\text{“E1”}; \theta_i^{(C)}\big), \tag{4}$$

where $\gamma_{j(\text{E1})}^{(C)}$ captures the attractiveness of response category E1 for item j under intuitive reasoning.

The expressions for Models A and B follow directly from the product of the branch probabilities in Figures 2 and 3. Both the CMR model and the extended versions of the CMR model are estimated by maximum marginal likelihood methods in combination with Gauss–Hermite quadrature (Bock & Aitkin, 1981). In the reported application, a quasi-Newton method is used that approximates the inverse Hessian according to the Broyden–Fletcher–Goldfarb–Shanno update (see Gill, Murray, & Wright, 1981). The algorithm utilizes the partial derivatives of the log-likelihood function with respect to all parameters and estimates the Hessian in the form of the cross-product of the Jacobian of the gradient. Model fits are assessed by standard methods. Because of the small number of items, we make use of the likelihood-ratio, goodness-of-fit test comparing observed and expected frequencies of the response patterns.

4. Analysis of Response Times

The dualistic model distinction between judgments that are formed easily and quickly as opposed to slowly and extensively provides one avenue for validating the categorization of E1, E2, and R responses. If E1 responses are a result of intuitive reasoning, they should be faster than either E2 or R responses, which are assumed to result from a more deliberate process. To test this conjecture, the response times for each of the five items were recorded for a subset of the respondents ($n = 117$). These response times were log-transformed and analyzed with a two-level regression model. For the analysis, both the items and the three response types were dummy coded, with the "speed" item and response R serving as reference categories. The resulting estimates are listed in Table 2. We note that response times differ systematically among items but that neither the effects of gender nor of match are significant. More important, E1 responses are significantly faster than E2 or R responses, whereas time differences between E2 and R responses are non-significant. These results provide additional support for the assignment of E1 and E2 responses to the two different stages in the CMR model.

TABLE 2.
Fixed-effects estimates of log response-time data.

Effect	Estimate	SE	p-val.
β_0	3.74	0.11	0.01
β_{female}	0.07	0.11	0.57
β_{match}	−0.0	0.12	0.43
$\beta_{\text{female}\times\text{match}}$	−0.02	0.16	0.91
β_{batball}	−0.21	0.07	0.01
β_{widgets}	−0.47	0.07	0.01
β_{lilypads}	−0.37	0.07	0.01
β_{married}	−0.21	0.07	0.01
β_{E1}	−0.29	0.06	0.01
β_{E2}	0.13	0.10	0.19

5. Model Comparisons

An attractive feature of the CMR models is that it allows testing whether the effects of covariates vary across stages. Specifically, in the reported application, it is of interest to determine whether a peak match increases inhibitory control as would be expected based on the Bodenhausen (1990) study, and whether this effect applies also to the deliberation stage by increasing the number of correct responses. Similarly, the observed gender effects may be caused by differences in inhibitory control, deliberate reasoning, or both factors.

To answer these questions, four different versions of the two-dimensional CMR model were fit to the data. The first model was the most general one and estimated the five location parameters, as well as the effect of peak match and the gender effect for both the intuitive and the reflective response part of the model. In addition, it allowed for a different covariance matrix for each of the four peak match and gender combinations. Thus, we estimated for the inhibitory-control stage and the deliberate response stages the two regression relations $\theta_i^{(C)} = PM_i\beta_1^{(C)} + G_i\beta_2^{(C)} + \epsilon_i^{(C)}$ and $\theta_i^{(D|C)} = PM_i\beta_1^{(D|C)} + G_i\beta_2^{(D|C)} + \epsilon_i^{(D|C)}$, respectively. This model yielded a log-likelihood of $-2{,}161.8$ with 26 parameters (i.e., 10 item parameters, 12 covariance terms and four regression coefficients) and an overall goodness-of-fit statistic of $G^2 = 411$ with $df = 942$. Although the latter fit statistic is of little value because the data matrix is rather sparse, detailed residual analyses comparing observed and expected frequencies for all item pairs and triplets demonstrated that this model provided a satisfactory fit of the data.

Subsequently, special cases of this general model were fitted to the data. A model with the same covariance matrix $\mathbf{\Sigma}$ for all four peak match and gender combinations demonstrated that there were no significant differences in the covariance structures across conditions ($\Delta G^2 = 8.0, df = 9$). In view of the sample size, we note that this test may have little power. Although peak match and gender did not predict variations in the reflective individual-difference factor $\theta_i^{(D|C)}$ ($\Delta G^2 = 1.4, df = 2$), they predicted variations in the inhibitory-control factor $\theta_i^{(C)}$ ($\Delta G^2 = 56.1, df = 2$). The log-likelihood of this CMR model version is $-2{,}166.4$ with 15 parameters (see Table 4).

As an additional test of the model specification, we also fit the extended versions of the CMR model depicted in Figures 2 and 3. Model A allows for misclassifications of E1 and E2 responses at both the intuitive and deliberate response levels. However, this model yielded almost the same fit as the CMR model, with a log-likelihood of $-2{,}166.0$ and both misclassification probabilities approaching their boundary values. Model B requires estimating an additional five location parameters as well as the gender and match effect in order to capture individual differences in E2 responses at the intuitive level; it yielded a log-likelihood of $-2{,}162.3$ with 22 parameters.

TABLE 3.
Orlando–Thissen χ^2-fit statistics of cognitive-miser response model.

Item	Male	Match	Male	NMatch	Female	Match	Female	NMatch
	1(23)	(12)3	1(23)	(12)3	1(23)	(12)3	1(23)	(12)3
batball	4.0	5.2	4.2	3.6	8.3	2.0	9.1	9.8
widget	3.8	3.6	3.8	4.4	5.8	4.1	6.4	7.9
lilypads	6.6	4.4	11.6	4.7	6.1	1.6	8.4	5.5
married	2.5	4.0	9.3	7.0	10.7	5.7	11.8	9.6
speed	9.1	4.1	4.2	11.6	13.6	5.1	7.4	6.9

TABLE 4.
Estimates of cognitive-miser response model.

Effects	Estimates	SE	Est/SE
$\gamma^{(C)}_{\text{batball}}$	−0.10	0.13	−0.73
$\gamma^{(C)}_{\text{widgets}}$	1.36	0.14	9.88
$\gamma^{(C)}_{\text{lilypads}}$	1.69	0.14	11.82
$\gamma^{(C)}_{\text{married}}$	0.01	0.12	0.05
$\gamma^{(C)}_{\text{speed}}$	−1.56	0.16	−9.82
$\gamma^{(D\mid C)}_{\text{batball}}$	3.10	0.37	8.49
$\gamma^{(D\mid C)}_{\text{widgets}}$	2.37	0.25	9.54
$\gamma^{(D\mid C)}_{\text{lilypads}}$	1.99	0.22	9.17
$\gamma^{(D\mid C)}_{\text{married}}$	1.78	0.27	6.64
$\gamma^{(D\mid C)}_{\text{speed}}$	1.43	0.37	−3.82
$\beta^{(C)}_{\text{match}}$	0.73	0.12	6.15
$\beta^{(C)}_{\text{female}}$	−0.60	0.12	−5.10
σ^2_D	2.91	0.76	3.82
σ_{CD}	0.51	0.22	2.34
σ^2_C	0.93	0.16	5.85

Although the CMR model and model B are not nested, it is clear from the small difference in the overall fit of both models that the extended version is not preferable. This conclusion was corroborated further by residual analyses of the two-dimensional CMR model. Table 3 reports the Orlando–Thissen χ^2-fit statistics (Orlando & Thissen, 2000) for each item by grouping response categories 2 and 3 (denoted by 1(23) in Table 3), as well as response categories 1 and 2 (denoted by (12)3 in Table 3). When compared to a χ^2-distribution with four degrees of freedom, none of the items appears to exhibit systematic misfit across the four conditions.

The parameter estimates of the CMR model are presented in Table 4. The estimated location parameters show that the ordering of the items is rather different for the deliberate and intuitive processes. The speed, married, and batball items trigger a high percentage of impulsive incorrect responses. The widgets and lilypads items give rise to fewer intuitive errors, but deliberate reasoning does not guarantee the correct response. Inhibitory control becomes less likely when the items are given at a time at which test takers are presumably less alert. It is interesting, however, that the reduction in inhibitory control does not seem to carry over to the execution of deliberate reasoning. Although test takers are more likely to give an impulsive incorrect response when they are less alert, being tired does not appear to substantially increase the probability of making a deliberate reasoning error for the considered test items. Female test takers appear to exhibit less inhibitory control in their answers, as shown by the negative regression effect. However, they do not differ significantly from male respondents in their deliberate-processing performance. Individual differences in inhibitory control are smaller, with an estimated variance of 0.93 (0.16), than differences in deliberate reasoning, with an estimated variance of 2.91 (0.76). The correlation between both individual-difference factors is positive, suggesting that respondents with higher inhibitory control are also more analytical processors, which increases their propensity to give a correct answer.

Overall, these results demonstrate the potential of the CMR model for analysis of dual-process data. Support for the proposed model structure was obtained in three ways. First, the extended versions of the CMR model did not fit better, which shows that it is unlikely that E1 and E2 responses are misclassified and that respondents arrived at E2 responses based on intuitive reasoning. Second, the two-dimensional CMR model provided a satisfactory fit for the data. Third, we obtained stage–specific effects of the covariates that were consistent with our hypotheses. The finding that a mismatch in peak time decreases inhibitory control agrees with the Bodenhausen (1990) study, which observed an increase in social stereotypic biases when judgments were rendered at a non-optimal time of the day. However, the result that a mismatch in peak time does not necessarily lead to more deliberate reasoning errors is new. This result suggests that other process characteristics, aside from response errors, may be needed in characterizing how System 1 and 2 performances are affected by external factors. We note that the response-time analysis did not point to systematic time differences between test takers whose peak time was matched or mismatched. Thus, being tired did not increase the error rate of the deliberate system and, apparently, did not slow it down.

Systematic sex differences were noted previously by Frederick (2005), who observed that female test takers tend to give more E1 responses. However, the present study also shows that these results hold on the item level, with each item differing in the degree to which it triggers intuitive and reflective answers. The finding that respondents differ more in their deliberate than in their inhibitory control suggests that automatic response processes may be less prone to individual variation than analytical processing. Further studies with additional items are desirable to determine the generalizability of this result. The positive correlation between individual differences in suppressing System 1 responses and individual differences in deliberate reasoning invites the search for experimental conditions that may weaken or even reverse this relationship.

6. A Rasch Analysis

The obtained results are critically dependent on the separate coding of the two *incorrect* response categories. When collapsing these two response categories and fitting a one-dimensional Rasch model to the binary data, we obtain a satisfactory fit ($G^2 = 92.6, df = 116$), indicating that the recoded data contain little information about the two individual-difference components. As in the previous analyses, the estimated effects of the two covariates, peak match and gender ($\hat{\beta}_{\text{match}} = 0.53(0.13)$, $\hat{\beta}_{\text{female}} = -0.46(0.13)$), show that females perform more poorly than male respondents and that a mismatch in time lowers performance. However, we note that instead of affecting impulsive response tendencies, in this analysis the covariates predict variations in ability levels. As a result, the effects of these factors are weaker because the two types of incorrect response are merged into one category, which dilutes their separate effects.

7. Concluding Remarks

This paper started with the question of whether it is possible to disentangle System 1 and System 2 sources of response tendencies on the basis of item responses alone. The presented application and the CMR model suggest that the answer is affirmative, provided certain conditions are satisfied. First, it is critical that an item or vignette gives rise to an answer that allows identifying the presence of an impulsive response process. Second, it is also important that suppressing the immediate answer does not by default give rise to the correct one but requires deliberative efforts that may lead to an incorrect response. If these two conditions are satisfied, the two-stage CMR model facilitates rigorous analyses of items and experimental conditions designed to measure System 1 and System 2 processing.

It is worth emphasizing that the correctness of an item response by itself is not sufficient to diagnose a deliberate or an intuitive process. For example, some items may have a high likelihood of being answered incorrectly based on deliberate processes but are solved correctly when intuitive processes are relied on, whereas for other items this relationship may be reversed. The presented study focused on items that were assumed to trigger intuitive incorrect responses but required some deliberation to be solved correctly. The response-time analysis, the time-of-day effect, and previous investigations utilizing these items (Alter, Oppenheimer, Epley, & Eyre, 2007; Frederick, 2005; Stanovich & West, 2008) provided support for this assumption. When such validating information is not available, it is critical to model the different types of incorrect response in combination with process-related indicators to determine the underlying response processes. For example, one promising avenue is to include response times in the modeling framework. Our response-time analysis showed that E1 responses were faster on average than E2 and R responses in the presented application. This result suggests that response times could prove valuable as indicators of System 1 and System 2 processing. Specifically, a two-class mixture model could be fit to the item-response times to distinguish between fast and slow item responses. Based on these classification results, item responses could be assigned to one of the two stages and analyzed with the CMR model.

Future research using models to test dual-process theories should take into account the potentially multidimensional nature of test items, multiple error categories, more complex relations between System 1 and System 2 processes, and/or additional process-related characteristics such as response times or confidence measures. Ultimately, the results of such studies could make separating the influence of cognitive abilities and other components of judgments and decision-making such as rationality or quality of thinking (Stanovich, 2009; Sternberg, 2000) a routine psychometric problem.

References

Alter, A., Oppenheimer, D., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*, 569–576.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29–51.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.

Bodenhausen, G.V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, *1*, 319–322.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models*. New York: Springer.

De Neys, W. (2006). Dual processing in reasoning. *Psychological Science*, *17*, 428–433.

Evans, J.S.B.T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42.

Gill, P., Murray, W., & Wright, M. (1981). *Practical optimization*. London: Academic Press.

Horne, J., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*, 97–110.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgments* (pp. 49–81). Cambridge: Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K.J. Holyoak & R.G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge: Cambridge University Press.

Logan, G.D., Schachar, R.J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science*, *8*, 60–64.

Orlando, M., & Thissen, D. (2000). New item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*, 50–64.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
Stanovich, K.E. (2009). *What intelligence tests miss: The psychology of rational thought*. New Haven: Yale University Press.
Stanovich, K.E., & West, R.F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*, 672–695.
Sternberg, R. (2000). *Handbook of intelligence*. Cambridge: Cambridge University Press.
Strack, F., Martin, L.L., & Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology*, *8*, 429–442.
Suh, Y., & Bolt, D.M. (2010). Nested logit models for multiple-choice item response data. *Psychometrika*, *75*, 454–473.
Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika*, *49*, 501–519.
Thurstone, L.L. (1927). *The nature of intelligence*. New York: Harcourt.
Tutz, F. (1997). Sequential models for ordered responses. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 139–152). New York: Springer.
Verhelst, N.D., Glass, C.A.W., & de Vries, H.H. (1997). A steps model to analyze partial credit. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 123–138). New York: Springer.

Manuscript Received: 30 AUG 2010
Final Version Received: 7 FEB 2011