T E C H N I C A L
R E P O R T

0251

INTERPRETATION OF THE DISCRIMINATION PARAMETER

TUERLINCKX, F. and P. DE BOECK

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Running head: INTERPRETATION OF THE DISCRIMINATION PARAMETER

Two interpretations of the discrimination parameter

Francis Tuerlinckx and Paul De Boeck

Department of Psychology, University of Leuven

Abstract

In this paper we propose two interpretations for the discrimination parameter in the Two-Parameter Logistic Model (2PLM). The interpretations are based on the relation between the 2PLM and two stochastic models. In the first interpretation, the 2PLM is linked to a diffusion model so that the probability of absorption equals the 2PLM. The discrimination parameter is the distance between the two absorbing boundaries and therefore the amount of information that has to be collected before a response to an item can be given. For the second interpretation, the 2PLM is connected to a specific type of race model. In the race model, the discrimination parameter is inversely related to the dependency of the information used in the decision process. Extended versions of both models with person-to-person variability in the difficulty parameter are considered. When fitted to a data set, it is shown that a generalization of the race model that allows for dependency between choices and response times (RTs) is the best-fitting model.

Keywords: Item Response Theory, Item Discrimination, Two-Parameter Logistic Model, Diffusion Model, Race Model, Response Times

Two interpretations of the discrimination parameter

In this paper we would like to shed some light on the interpretation of the discrimination parameter in the Two-Parameter Logistic Model (2PLM; Birnbaum, 1968). Assume a person responds to an item with two response alternatives, A and B. Let X be a random variable that takes the value 1 if alternative A is chosen and 0 if the person chooses alternative B. A possible model for the choice probability is the 2PLM (Birnbaum, 1968):

$$\Pr(X = 1) = \frac{\exp\left(\alpha(\theta + \beta)\right)}{1 + \exp\left(\alpha(\theta + \beta)\right)}, \tag{1}$$

where the parameters θ and β are the location parameters because they refer to the positions on the latent continuum of the person and item, respectively. The parameter α is known as the item discrimination parameter. In the basic interpretations, the item discrimination quantifies how well the single latent dimension is measured by the item. Likewise, it is a measure of the nonlinear correlation between the item and the latent dimension. Such interpretations resemble the interpretation of a factor loading in factor analysis. However, both accounts are merely verbal paraphrases of the role of α in Equation (1). More explanatory interpretations have been proposed by Colonius (1981) and, more recently, by Embretson (1999).

In the following, two new interpretations of the discrimination parameter will be developed by considering the 2PLM to be the choice probability arising from two different stochastic models: a diffusion model and a race model. For both type of models, we will indicate under which conditions the choice probabilities equal the 2PLM and how the discrimination parameter may be interpreted.

The organization of the rest of the paper is as follows. First, the diffusion model is discussed and then the race model. It is shown that both models assume independence of choices and RTs. Extensions of both models are proposed that relax the independence assumption. Subsequently, we present an example in which the models are fitted to personality inventory data. Finally, the paper is closed with a discussion of the results.

The diffusion model and the 2PLM

In the diffusion model it is assumed that in order to decide between alternatives A and B, information has to be sequentially sampled from the stimulus and mapped onto a single signed counter. If the counter hits an upper or lower boundary, the corresponding response is given (i.e., alternative A for the upper boundary and B for the lower one). A Wiener process with constant drift and variance and two absorbing boundaries (Cox & Miller, 1970) can represent such an accumulation process mathematically. Throughout the rest of the paper, we will use the terms "diffusion process" or "diffusion model" to refer to this Wiener process, although many other diffusion processes exist (e.g., Cox & Miller, 1970).

The amount of accumulated information at time d is denoted by the continuous random variable $Z(d)$. At the start of the deliberation process (i.e., $d = 0$), $Z(0)$ equals z. The process ends if the evidence counter reaches an upper absorbing boundary α or a lower absorbing boundary 0 at some time (where $0 < z < \alpha$). The time needed to reach one of the boundaries is represented by a random variable D, which is the decision time.

There are three other essential characteristics of the diffusion process. First, the mean rate of information accumulation of the process is called the drift rate, denoted by μ, and it is assumed to be constant over time. If μ is positive and large, this means that a person tends to sample more information in favor of the alternative corresponding to the upper boundary (alternative A). If it is large but negative, the person tends to sample more evidence for alternative B. It will be assumed that the drift can be decomposed as follows:

$$\mu = \theta + \beta, \tag{2}$$

where θ is the person component and β the item (or stimulus) component of the drift rate.

Holding θ constant, information in favor of the upper barrier alternative can be found more easily for an item with a larger β compared to one with a smaller β. Hence, the parameter β is an item main effect on the drift rate. Ratcliff (1985) and Ratcliff, Van Zandt, and McKoon (1999) present a related idea to the decomposition in Equation (2) where the $-\theta$ is a criterion threshold that has to be exceeded by β to attain a positive drift rate. In our model, these criterion thresholds are subject to interindividual differences.

A second feature of the information accumulation process is its volatility. While the mean rate of information accumulation is constant over time, the actual accumulation rate varies over time due to the stochastic nature of the process. In a small time interval Δd, the variance of the change in the internal counter $Z(d)$ equals $\sigma^2 \Delta d$. If this variance is large, the process behaves more volatile, and if it is small, the trajectory of $Z(d)$ over time will not deviate much from the mean drift rate.

A third property of the diffusion process is the location of the starting point z. Dependent on the value for the starting point, there may be a bias toward one of the response alternatives. If $z > \alpha/2$, there is a bias toward alternative A, while $z < \alpha/2$ indicates a bias toward alternative B. In the case that $z = \alpha/2$, there is no bias.

The observed response time (RT) will not be exactly equal to the decision time because additional processes take place in the time between the onset of the stimulus and the final response besides the decision. Therefore, the observed RT, represented by the random variable T, is decomposed into the decision time D and a residual time T_{er}:

$T = D + T_{er}$.

The residual time represents for instance the time needed to encode the item and to prepare the response. For simplicity, it is assumed that T_{er} is constant over persons and items. For an extension of the model where the residual time is assumed to be randomly distributed, see Ratcliff and Tuerlinckx (2002).

Let the random variable X denote the choice response (i.e., the boundary of absorption) and the random variable T the observed RT. The joint density of X and T can then be written as follows (Cox & Miller, 1970):

$$\begin{aligned} f_{X,T}(x,t) &= \frac{\pi\sigma^2}{\alpha^2}\exp\left(\frac{(\alpha x - z)(\theta+\beta)}{\sigma^2} - \frac{(\theta+\beta)^2}{2\sigma^2}(t - T_{er})\right) \\ &\times \sum_{m=1}^{\infty} m \sin\left(\frac{\pi m(\alpha x - 2zx + z)}{\alpha}\right)\exp\left(-\frac{1}{2}\frac{\pi^2\sigma^2 m^2}{\alpha^2}(t - T_{er})\right), \end{aligned} \tag{3}$$

which is only nonzero for $t > T_{er}$. Because the joint density in Equation (3) does not change if the parameters $\theta, \beta, \alpha, \sigma$ and z are all multiplied by a constant K, the model is not identified and therefore one of the parameters has to be restricted to an arbitrary value. For simplicity, we set $\sigma = 1$.

From Equation (3), the probability of absorption at the upper boundary (see e.g., Cox & Miller, 1970; Luce, 1986): Pr(choosing alternative A) $= \Pr(X = 1) = \frac{\exp(-2z(\theta+\beta))-1}{\exp(-2\alpha(\theta+\beta))-1}$. If the diffusion process is unbiased ($z = \alpha/2$), then the probability of absorption in the upper boundary reduces to:

$$\Pr(X = 1) = \frac{\exp\left(-\alpha(\theta+\beta)\right) - 1}{\exp\left(-2\alpha(\theta+\beta)\right) - 1} = \frac{\exp\left(\alpha(\theta+\beta)\right)}{1 + \exp\left(\alpha(\theta+\beta)\right)}, \tag{4}$$

which shows that the choice probability model for an unbiased diffusion model with absorbing barriers and with drift decomposed into a person and an item part equals the 2PLM. In the unbiased diffusion process, the discrimination parameter α is the distance between the two absorbing boundaries.

The diffusion model interpretation of the item discrimination implies that items with different discrimination parameters have different speed-accuracy trade-off functions (SATFs) (Luce, 1986), as if they were different conditions in an experiment. The SATF refers to the relation between the expected RT and the "accuracy" of the response. A response is defined to be accurate in this context if the mean drift rate $\mu = \theta + \beta$ and the manifest response (X) match. Such a match occurs if the case $\theta + \beta > 0$ corresponds with reaching the upper boundary and therefore response alternative A is chosen ($X = 1$). (Vice versa, choosing response alternative B ($X = 0$) when $\theta + \beta > 0$ results in a mismatch.) A match also occurs if $\theta + \beta < 0$ and $X = 0$. From Equation (4), it can be seen that the distance between the boundaries (or item discrimination) is related to the probability of misclassification. When α increases, $\Pr(X = 1)$ increases if $\theta + \beta > 0$, so that the probability of giving a response that is in correspondence with the latent disposition increases. In addition, a larger α results in a larger expected decision time (since more information has to be accumulated, and that takes a longer time), but the response is expected to be more accurate. In conclusion, better discriminating items will require more time under the diffusion model.

Usually, the boundary separation is seen as a criterion under subject control while in our interpretation it is a property of an item. However, this is not a contradiction. In the model proposed here, it is possible that the subjects determine the location of the absorbing boundaries, but that the item has an influence on the subjects in setting the boundaries (see also Grice, 1968); therefore, it can be seen as an item main effect.

The race model and the 2PLM

In the race model described here, it is assumed that each choice alternative (A or B) is represented by a decision node. Each decision node receives connections from several information nodes, as shown in Figure 1. After encoding the item, the information nodes will become activated and they consequently perform a race to activate their corresponding decision node. The first-activated decision node determines the choice response.

INSERT FIGURE 1 ABOUT HERE

The number of information nodes associated with each decision node, denoted as N_A and N_B, is determined by the person and item parameter as follows:

$$\begin{aligned} N_A &\approx N\exp(\theta)\exp(\beta) \\ N_B &\approx N\exp(-\theta)\exp(-\beta), \end{aligned} \tag{5}$$

where N is the grand mean of the number of information nodes and $\exp(\theta)$, $\exp(\beta)$ and their inverses represent deviations from this grand mean. The number of information nodes is restricted to be positive, therefore they are decomposed in a multiplicative way, resembling a loglinear representation for a frequency table (Agresti, 2002). The left-hand and right-hand sides in Equation (5) are only approximately equal because the number of actual information nodes has to be a natural number and $\exp(\theta)$ and $\exp(\beta)$ are usually not natural numbers. In the following, however, we will assume that the equality holds.

Next, we consider the (marginal) race time to pass activation from the ith information node to decision node u ($u = A, B; i = 1, \ldots, N_u$), denoted as T_{ui}. It is assumed that this race time is Weibull distributed with scale parameter λ and shape parameter γ so that its survivor function is:

$$S_{T_{ui}}(t) = \Pr(T_{ui} > t) = \exp\left(-\lambda t^{\gamma}\right). \tag{6}$$

The mean of this distribution is $\lambda^{-\frac{1}{\gamma}}\Gamma\left(1+\frac{1}{\gamma}\right)$ (with $\Gamma(\cdot)$ is the gamma function). If γ equals 1, then Equation (6) becomes the survivor function of an exponential distribution.

The univariate marginal race times for the information nodes are assumed to be distributed identically, but the race times of information nodes connected to the same decision node will not have to be independent. We allow for dependencies in race times because the information nodes may coactivate each other in the decision process, leading to positively correlated race times.

To induce the within-decision node dependency between the race times, we use the technique of copulas (Joe, 1997). A copula is a function used to create a joint survivor function from given univariate marginal survivor functions with a certain dependency structure, while still retaining the Weibull survivor functions from Equation (6) as the univariate marginals. As suggested already by the term copula, several univariate marginal survivor functions are "coupled" to yield a joint survivor function, but without affecting the univariate marginal survivor functions.

In this paper, we use a multivariate generalization of a Gumbel-Hougaard copula (Hougaard, 1986a, 1986b; Joe, 1997) applied to the marginal Weibull survivor functions from Equation (6). The joint survivor function of the vector of race times connected to decision node u, represented as $(T_{u1}, \ldots, T_{uN_u})$, then equals:

$$\begin{aligned} S_{T_{u1},\ldots,T_{uN_u}}(t_{u1},\ldots,t_{uN_u}) &= \Pr(T_{u1} > t_{u1},\ldots,T_{uN_u} > t_{uN_u}) \\ &= \exp\left[-\left(\sum_{i=1}^{N_u} \lambda^{\omega} t_{ui}^{\gamma\omega}\right)^{\frac{1}{\omega}}\right], \end{aligned} \tag{7}$$

where $u = A, B$ and $i = 1, \ldots, N_u$ and ω is a dependency parameter (with $\omega \geq 1$ to ensure a proper probability density). It is easy to see that the marginal survivor function of T_{ui} is still equal to Equation (6):

$$\begin{aligned}\Pr(T_{ui} > t) &= \Pr(T_{u1} > 0, \ldots, T_{ui} > t, \ldots, T_{uN_u} > 0) \\ &= S_{T_{u1},\ldots,T_{uN_u}}(0, \ldots, t_{ui}, \ldots, 0) = \exp\left(-\lambda t^{\gamma}\right). \end{aligned} \tag{8}$$

The parameter ω in Equation (7) is the dependency parameter and it quantifies the amount of dependency between the race times of the information nodes connected to a single decision node. The case $\omega = 1$ refers to independence because then the multivariate survivor function in Equation (7) becomes the simple product of the univariate ones as given in Equation (6). If ω increases, the dependency between the race times also increases. As $\omega \to \infty$ (perfect dependency between the race times), the joint survivor function becomes equal to $\exp\left[-\max_i\left(\lambda t_{ui}^{\gamma}\right)\right] = \min_i \exp\left(-\lambda t_{ui}^{\gamma}\right)$, implying that the joint survivor function only depends on the fastest race time.

In a next step, the distribution of $T_A = \min(T_{A1}, \ldots, T_{AN_A})$, the minimum of the race times from information nodes connected to decision node A, is derived:

$$\begin{aligned}S_{T_A}(t) &= \Pr(T_A > t) = \Pr(\min(T_{A1}, \ldots, T_{AN_A}) > t) = \Pr(T_{A1} > t, \ldots, T_{AN_A} > t) \\ &= S_{T_{A1},\ldots,T_{AN_A}}(t, \ldots, t) = \exp\left(-\lambda N_A^{\frac{1}{\omega}} t^{\gamma}\right). \end{aligned} \tag{9}$$

As can be seen from Equation (9), T_A is distributed as a Weibull random variable with parameters $\lambda N_A^{\frac{1}{\omega}}$ and γ. A similar derivation leads to the survivor function of T_B, which is Weibull distributed with parameters $\lambda N_B^{\frac{1}{\omega}}$ and γ. The two random variables T_A and T_B are independently distributed.

With the distribution of random variables T_A and T_B, we are in a position to define the choice probabilities and the distribution of the decision time $D = \min(T_A, T_B)$. We start with the probability of choosing alternative A:

$$\begin{aligned}\Pr(X=1) &= \Pr(T_A < T_B) = \int_0^{\infty} f_{T_A}(t_A)\left(\int_{t_A}^{\infty} f_{T_B}(t_B)dt_B\right)dt_A \\ &= \frac{N_A^{\frac{1}{\omega}}}{N_A^{\frac{1}{\omega}} + N_B^{\frac{1}{\omega}}} = \frac{\exp\left(\frac{2}{\omega}(\theta+\beta)\right)}{1+\exp\left(\frac{2}{\omega}(\theta+\beta)\right)} = \frac{\exp\left(\alpha(\theta+\beta)\right)}{1+\exp\left(\alpha(\theta+\beta)\right)}, \end{aligned} \qquad (10)$$

where we have used Equation (5) and set $2/\omega$ equal to α.

From Equation (10), we see that under the race model the discrimination parameter α equals $2/\omega$, meaning that it is a measure of dependency between the information nodes. Therefore, the larger ω, the larger dependency there is among the information nodes, and the smaller α will be. Conversely, the closer ω to 1, the smaller the dependency among the information nodes and therefore the larger α. Consequently, α is bounded between 0 and 2 but neither the lower limit nor the upper limit leads to problems: The former is natural and the latter is only a matter of convention.

In the diffusion model, items with different discriminations had different SATFs. The same holds for the race model. If $\theta + \beta > 0$, then one can deduce from Equation (5) that this means that $N_A > N_B$. A correct classification occurs in this case if alternative A is chosen and a misclassification if alternative B is chosen. The discrimination parameter α influences the probability of a misclassification in the same way as in the diffusion model (if α increases, so does $\Pr(X = 1)$ when $\theta + \beta > 0$). But, if α increases, the faster the response is expected because the information nodes behave more independently. Dependency among the information nodes can be understood as if a smaller number of effective information nodes were involved in the decision. In the extreme case of perfect dependency, the race would consist of one node against another one because all information node race times within a decision node are equal. To illustrate this, let $\omega \to \infty$, then we can see that the survivor function of the minimum of the race times for decision node A in Equation (9) becomes equal to $\exp\left(-\lambda t^{\gamma}\right)$ (because $N_A^{\frac{1}{\omega}} \to 1$ if $\omega \to \infty$). However, this is the survivor function of the race time of a single information node involved in the race to activate decision node A. Therefore, less dependency means that more nodes are effectively participating in the race such that the probability of a faster ending becomes more likely. Hence, opposite to the diffusion model, in the race model we expect that strongly discriminating items take less time to respond.

As for the diffusion model, it is reasonable to decompose the total RT T into the sum of the decision time and some residual time, the latter denoted again as T_{er}. The decision time is represented by a random variable D and equals $\min(T_A, T_B)$. Thus, the survivor function for T can be derived as follows:

$$\begin{aligned}
S_T(t) &= \Pr(T_A + T_{er} > t, T_B + T_{er} > t) = \Pr(T_A > t - T_{er})\Pr(T_B > t - T_{er}) \\
&= \exp\left(-\lambda\left[N_A^{\frac{1}{\omega}} + N_B^{\frac{1}{\omega}}\right](t - T_{er})^{\gamma}\right) \\
&= \exp\left(-\lambda N^{\frac{1}{\omega}}\left[\exp\left(\frac{1}{\omega}(\theta+\beta)\right) + \exp\left(-\frac{1}{\omega}(\theta+\beta)\right)\right](t - T_{er})^{\gamma}\right) \\
&= \exp\left(-\lambda N^{\frac{\alpha}{2}}\left[\exp\left(\frac{\alpha}{2}(\theta+\beta)\right) + \exp\left(-\frac{\alpha}{2}(\theta+\beta)\right)\right](t - T_{er})^{\gamma}\right) \\
&= \exp\left(-2\lambda N^{\frac{\alpha}{2}}\cosh\left(\frac{\alpha}{2}(\theta+\beta)\right)(t - T_{er})^{\gamma}\right),
\end{aligned} \tag{11}$$

where the last expression follows from the definition of the hyperbolic cosine function: $\cosh(x) = \frac{1}{2}\left(e^{x} + e^{-x}\right)$. Note that the survivor function of the total RT in Equation (11) is the survivor function of a Weibull random variable with a support shifted away from zero by an amount T_{er}.

From the derivation of Equation (11), it follows that that the choice and RT are independent in the race model, which is a well-known property of the minimum of two Weibull distributions (see e.g., Joe, 1997). The density $f_T(t)$ corresponding to the survivor function $S_T(t)$ can be obtained easily: $f_T(t) = -\frac{dS_T(t)}{dt}$. Multiplying the RT density with the probability from Equation (10) gives the joint density of X and T:

$$\begin{aligned} f_{X,T}(x,t) &= 2\lambda\gamma N^{\frac{\alpha}{2}} \frac{\exp\left(x\alpha(\theta+\beta)\right)}{1+\exp\left(\alpha(\theta+\beta)\right)} \cosh\left(\frac{\alpha}{2}(\theta+\beta)\right)(t-T_{er})^{\gamma-1} \\ & \quad \times \exp\left(-2\lambda N^{\frac{\alpha}{2}} \cosh\left(\frac{\alpha}{2}(\theta+\beta)\right)(t-T_{er})^{\gamma}\right) \end{aligned} \tag{12}$$

Compared to the diffusion model, the race model has three additional parameters: the shape parameter γ, the connection strength λ and the grand mean of the number of information nodes N. However, not all parameters in the race model are identified. If we multiply θ and β with a constant K, divide α by the same constant K, and raise N to the power K, then the joint density of X and T does not change. This indetermination can be solved by fixing N to a constant; we choose $N = 10$.

The conditional accuracy function (CAF)

The diffusion model and the race model discussed in this paper have the strong assumption that given the parameter values, the choice and RT are independent. For the race model, the independence of choice and RT has already been mentioned above. For the diffusion model, the independence assumption can be verified by inserting $z = \frac{\alpha}{2}$ into Equation (3) and deriving $f_{T|X}(t|X=0) = \frac{f_{X,T}(0,t)}{\Pr(X=0)}$ and $f_{T|X}(t|X=1) = \frac{f_{X,T}(1,t)}{\Pr(X=1)}$. It turns out that $f_{T|X}(t|X=0) = f_{T|X}(t|X=1) = f_T(t)$. Thus, in both models, the length of the decision time is not informative about the choice probability.

As a consequence of the independence assumption, the mean RTs of both choices, given the parameter values, are also equal. Therefore, if one would subtract the conditional mean RT of choosing option A from the conditional mean RT of choosing option B and plot it for several values of θ, the resulting function would be a horizontal line located at zero. This is also called a flat conditional accuracy function (CAF). The assumption of independence between choice and RT is a very restrictive one and often it is not justified by the data (see e.g., Ratcliff & Rouder, 1998; Ratcliff et al., 1999). Therefore, we discuss here for both models some possible extensions that do not rely on the independence between choice and RT.

A non-flat CAF in the diffusion model For the diffusion model, traditionally two modifications have been proposed to accommodate a non-flat CAF. The first modification is to allow for variability in some of the parameters. It has been shown (see e.g., Ratcliff & Rouder, 1998; Ratcliff et al., 1999) that additional random variability in drift and starting point from one item to the other can account for a non-flat CAF. Variability in starting point will not be considered further, because it leads to a functional form for the choice probabilities that deviates from the logistic 2PLM. For variability in drift however, the marginal choice probabilities are almost similar to the 2PLM as is demonstrated below.

Drift variability models interactions between persons and items. In Equation 2, the drift is a simple sum of a person and item parameter. However, under drift variability, the item parameter β differs over persons. Another interpretation, which amounts to the same, is that the person parameter θ varies over items. Thus, in case of random drift, Equation (2) becomes: $\mu = \theta + \beta + \epsilon$, with the common assumption: $\epsilon \sim \mathrm{N}(0, \tau^2)$. When ϵ is added to Equation (3), the integral over the normal density of ϵ has a closed-form solution for the joint probability density given that the process is unbiased ($z = \alpha/2$; see Tuerlinckx, 2004).

If the drift varies from item-to-item (for the same person), this induces dependency between the choice response and response time. Ratcliff and Rouder (1998) illustrate why this is the case. Let us suppose that the $\theta + \beta$ equals 2, but that the variability causes it to be 1.5 or 2.5 with equal probability. Assume that the item discrimination is 0.5, such that the probability of choosing alternative A (B) equals 0.8 (0.2) and 0.9 (0.1), respectively. Furthermore, assume that the mean response times of choosing alternative A

for both actual drift rates are 1.4s and 1.3s, respectively. Because of the independence property, the mean response times for choosing alternative B are also 1.4s and 1.3s. However, if we look at the mean response time of choosing A, averaged over the drift rate distribution, then this equals: $\frac{0.8}{0.8+0.9} \times 1.4\text{s} + \frac{0.9}{0.8+0.9} \times 1.3\text{s} \approx 1.3\text{s}$. However, for alternative B, the mean response time is: $\frac{0.2}{0.2+0.1} \times 1.4\text{s} + \frac{0.1}{0.2+0.1} \times 1.3\text{s} \approx 1.4\text{s}$. Thus, the least likely alternative (B in this case) will have the largest mean response time after averaging over the random drift distribution.

When considering the marginal probability of $X = 1$, we can make the following derivation:

$$\begin{aligned}
\Pr(X = 1) &= \int_{-\infty}^{\infty} \frac{\exp\left(\alpha(\theta + \beta + \epsilon)\right)}{1 + \exp\left(\alpha(\theta + \beta + \epsilon)\right)} \phi(\epsilon; 0, \tau^2) d\epsilon \\
&\approx \int_{-\infty}^{\infty} \Phi\left(\frac{\alpha}{1.7}(\theta + \beta + \epsilon)\right) \phi(\epsilon; 0, \tau^2) d\epsilon \\
&= \Phi\left(\frac{\alpha}{1.7\sqrt{\frac{\tau^2\alpha^2}{1.7^2} + 1}}(\theta + \beta)\right) \\
&\approx \frac{\exp\left(\frac{1.7\alpha}{\sqrt{\tau^2\alpha^2 + 1.7^2}}(\theta + \beta)\right)}{1 + \exp\left(\frac{1.7\alpha}{\sqrt{\tau^2\alpha^2 + 1.7^2}}(\theta + \beta)\right)},
\end{aligned}$$

where $\phi(\epsilon; 0, \tau^2)$ denotes the normal probability density with mean zero and standard deviation τ. In the derivation, we have made use twice of the approximate relationship: $\frac{\exp(x)}{1+\exp(x)} \approx \Phi\left(\frac{x}{1.7}\right)$. Lord and Novick (1968) show that the maximum absolute error of this approximation is smaller than 0.01. The integral on the second line of the derivation has a closed-form solution (see e.g., McCulloch & Searle, 2001, p229-230).

Hence, even with random drift, the marginal choice probabilities (after integrating out ϵ) are almost exactly equal to the 2PLM. The only difference is that marginally the new discrimination parameter is equal to $\frac{1.7\alpha}{\sqrt{\tau^2\alpha^2+1.7^2}}$ (which is only a monotone increasing transformation of α).

A second modification to allow for a non-flat CAF might be to consider other and more general types of diffusion processes than the Wiener process, such as the Ornstein-Uhlenbeck (OU) process in which the drift rate depends on the amount of already accumulated information (Busemeyer & Townsend, 1993; Diederich, 1997). However, the marginal choice probability is not equal anymore to the 2PLM and thus these options will not be considered anymore.

A non-flat CAF in the race model As for the diffusion model, it is possible to model in the race model interactions between persons and items by allowing for item-to-item random variation in the person parameter (or vice versa, person-to-person variation in the item parameter). Hence, we will also consider the random $\theta + \beta$ generalization for the race model[1]. (However for the race model, the integration over the probability density of ϵ has to be approximated numerically, as opposed to the diffusion model where a closed-form solution exists.)

Moreover, in the specific case of the race model, there is another very natural extension (for which there is no corresponding alternative in the diffusion model) that allows for dependency between choice and RT and that has an exact the connection with the 2PLM. For the choice probabilities, the 2PLM is taken as a starting point. Then, we make the conditional survivor function dependent on the choice response as follows:

$$S_{T|X}(t|x) = \exp\left(-2\lambda N^{\frac{\alpha}{2}} e^{x\delta} \cosh\left(\frac{\alpha}{2}(\theta+\beta)\right)(t - T_{er})^{\gamma}\right), \tag{13}$$

with the additional parameter δ that regulates the dependency of the RT on the choice response. If $\delta = 0$ the model simplifies to the model from Equation (11).

Using the formula for the mean of a Weibull distribution, we can derive an expression for the difference in conditional mean RTs in the delayed race model. For simplicity, let W denote $\left[2\lambda N^{\frac{\alpha}{2}} \cosh\left(\frac{\alpha}{2}(\theta+\beta)\right)\right]^{-\frac{1}{\gamma}} \Gamma\left(1 + \frac{1}{\gamma}\right)$. Then the difference between the two conditional RTs, $E(T|X = 0) - \mathrm{E}(T|X = 1)$, equals $W\left(1 - \exp\left(-\frac{\delta}{\gamma}\right)\right)$, where $W > 0$. If δ equals zero, there is no difference between the conditional mean RTs for $X = 0$ and $X = 1$. However, if $\delta < 0$, then $W\left(1 - \exp\left(-\frac{\delta}{\gamma}\right)\right) < 0$ and therefore $\mathrm{E}(T|X = 1)$ is larger than $\mathrm{E}(T|X = 0)$. In that case, choosing option A takes longer than choosing option B. The reverse relation follows for $\delta > 0$.

This variant will be called the "delayed" race model because it is assumed that the decision node for alternative A can be delayed (if $\delta < 0$) without altering the choice response. The magnitude of the delay depends on the value of δ. On average, the mean time to activate the decision node A is changed by a factor $e^{-\frac{\delta}{\gamma}}$. Of course, it is also possible that choosing option A is speeded up (which happens if $\delta > 0$), but that is equivalent to delaying option B. On top of the "delayed" extension of the race model, one could additionally allow for variability in $\theta + \beta$, as in the regular race model.

Example

Method and data In this section, we apply the models presented above to a data set taken from a study in which 96 second-year students had to fill out a computerized self-report personality questionnaire. Only a brief summary of the data collection method is presented here, more details can be found in Tuerlinckx and De Boeck (2004).

The materials were constructed by selecting unipolar traits from a variety of sources (English and Dutch Big Five markers and a set of traits generated by lay persons). All English trait names were translated to Dutch. For some of the traits, antonyms were already available because the traits were selected together with their antonyms from the bipolar trait pairs; however, for other traits, antonyms had to be collected. Eventually, we constructed from 97 unipolar traits, 61 bipolar trait pairs. Some traits were used two or three times, but no identical bipolar trait pairs were constructed.

Besides the 61 target trait pairs related to neuroticism and emotional stability, 102 filler bipolar trait pairs (related to the other four factors of the Big Five) were composed and mixed with the 61 target trait pairs. The neurotic alternative was randomly assigned to the left or right position and one questionnaire form was produced. The first eight items of the questionnaire were items from the other four factors and the first five of those served explicitly as examples to practice the instructions. The participants were tested in separate small cubicles equipped with a personal computer. The 17" screen was placed at eye level.

The 96 undergraduate participants were asked to choose the trait of the pair that applies the most to them. They were instructed explicitly to work as accurate and as fast as possible. They could make a choice by pressing one of two colored keys on a standard keyboard. Response time was the elapsed time from the appearance of the pair on the screen until the response was emitted. Stimuli were given in blocks of 10.

Only items with an item-test correlation larger than 0.3 were considered as a candidate for the analysis (for a justification of this threshold, see Ebel & Frisbie, 1986). From the reduced pool of 41 items, ten items were randomly sampled for the illustrative analysis in this paper (because the models were computationally quite complex only ten items were retained). Descriptions of these items can be found in Table 1 (for the trait pairs, we used the sans serif font and the neurotic alternative is printed in *italic*).

Let P denote the number of persons ($P = 96$) and I the number of items ($I = 10$). The entry in the data matrix pertaining to person p ($p = 1, \ldots, P$) and item i ($i = 1, \ldots, I$) contains two observations: a choice response x_{pi} and a RT t_{pi}. The pair of observations (x_{pi}, t_{pi}) is the realization of the pair of random variables (X_{pi}, T_{pi}) where X_{pi} takes the value 1 if person p chooses the neurotic alternative for item i and 0 if the person chooses the emotional stable one. The random variable T_{pi} refers to the time person p needed to respond to item i. Model parameters will also be indexed if necessary: θ becomes θ_p indicating that it is a person-specific parameter and α and β become α_i and β_i implying that they are specific to an item. The remaining parameters do not vary with persons nor items and thus do not carry an index. If the sum of θ_p and β_i is assumed to be random (as in the random drift diffusion model), then the quantity ϵ_{pi} is added.

INSERT TABLE 1 ABOUT HERE

Fitted models In order to obtain parameter estimates, we used the maximum marginal likelihood procedure (see e.g., Fischer & Molenaar, 1995). It is assumed that the θ_ps are a random sample from a normal population distribution with mean zero and standard deviation η. (The population mean is set equal to zero for reasons of identification.) Let $f_{X_{pi},T_{pi}}(x_{pi}, t_{pi})$ denote the density for a pair of observations. Marginalizing over the person-specific parameters θ_p gives the marginal likelihood, which will maximized with respect to the parameters after taking the logarithm:

$$\ell = \log \left(\prod_{p=1}^{P} \int_{-\infty}^{\infty} \prod_{i=1}^{I} f_{X_{pi},T_{pi}}(x_{pi}, t_{pi}) \phi(\theta_p; 0, \eta^2) d\theta_p \right), \tag{14}$$

where $\phi(\theta_p; 0, \eta^2)$ denotes the normal probability density with mean zero and standard deviation η. The integral in Equation (14) is intractable for the models discussed in this paper and it will be approximated by a Gauss-Hermite quadrature with 10 nodes (e.g., Naylor & Smith, 1982). All models discussed in this paper can be fitted with SAS PROC NLMIXED[2] (SAS Institute Inc., 1999).

An overview of the fitted models can be found in the second column of Table 2. For the diffusion and race models, we estimated the standard versions (see Equations (3) and (12)), a variant with equal item discriminations (second and fourth row in Table 2) and an extension of the model with random $\theta_p + \beta_i$. The delayed race model variant, as proposed in Equation (13), is also considered as well as a version with random $\theta_p + \beta_i$.

Fit of the 2PLM Because we are looking for an explanation of the discrimination parameter of the 2PLM, we need to make sure in the first place that the 2PLM fits the choice data. Parameter estimates under the 2PLM (based only on the choice data) are computed using the maximum marginal likelihood procedure outlined in the previous section. A difference is that because of identification problems when using only the choice data, we have to restrict the standard deviation of the distribution of θ_ps to a known constant (in this case it will be 1) so that all discrimination parameters are estimable.

To test whether the 2PLM fits the data, we conducted two goodness-of-fit tests based on a parametric bootstrap approach (Efron & Tibshirani, 1993). The parametric bootstrap method allows one to define any kind of test statistic and simulate (instead of deriving analytically) its reference distribution. As a first test, an omnibus test was chosen. Fitting the 2PLM to the observed choice data with the maximum likelihood method gives as a by-product the deviance, defined as $\text{dev} = -2\ell$. For the 2PLM applied to the observed data, this deviance was equal to 931. Next, the estimated parameters were used to simulate 100 replicated data sets according to the 2PLM. The 2PLM was fitted again to each simulated data set providing us with 100 replicated deviance values. The replicated deviances are then used to estimate the p-value of the test as follows:

$$\hat{p} = \frac{1}{100} \sum_{j=1}^{100} I(\text{dev}_j^{\text{rep}} > \text{dev}^{\text{obs}}), \tag{15}$$

where $\text{dev}_j^{\text{rep}}$ refers to the deviance obtained from the jth replicated data set and $I(C)$ is the indicator function taking value 1 if the condition C is true and zero otherwise. If the model fits the data, the observed deviance should not deviate too much from the simulated deviances, resulting in a $\hat{p}$-value that is not too extreme (not too small nor too large). We obtained $\hat{p} = 0.65$ which indicates a sufficient good fit of the 2PLM to the observed data.

A second test is a graphical one. In a first step, we computed the number of endorsed neurotic alternatives s_p by participant p (i.e., the sum score) and classified the s_ps in three sum score groups: Low ($s_p = 0, 1$), Middle ($s_p = 2, 3, 4$) and High ($s_p = 5, 6, 7, 8, 9, 10$). For the observed data, the frequencies in the three groups are 27, 44, and 25, respectively. Then we calculated for each item and each sum score group the proportion of persons endorsing that item. The crosses in Figure 2 are the observed proportions. Next, we simulated again 100 data sets using the parameter estimates from the observed data sets and repeated the aforementioned procedure for all 100 replicated data sets. The thick bars in Figure 2 denote the 95% confidence intervals of the simulated proportions. It can be seen that the observed proportions almost never fall outside the 95% confidence bounds. Based on the two previous tests, we conclude that the 2PLM fits the choice data sufficiently well.

INSERT FIGURE 2 ABOUT HERE

Model selection Following the estimation of the parameters for the several diffusion and race models, the model fit of each model has to be assessed using methods of model selection and model checking (for the latter, see the subsection below). For nested models, model selection can be accomplished using likelihood ratio tests. The likelihood ratio test statistic is computed by subtracting the deviance of the full model from the deviance from the reduced model. Asymptotically, the reference distribution for the test statistic (denoted as L) is a chi-square distribution with degrees of freedom the additional number of parameters of the full model compared to the reduced model.

Non-nested models (such as the diffusion model and the race model) must not be compared with the likelihood ratio test and model selection is then often based on

Akaike's information criterion (AIC; Akaike, 1977) and the Bayesian information criterion (BIC; Schwarz, 1978). Both measures add to the deviance of a model a penalty for the number of free parameters: For AIC the penalty equals $2k$ (with k the number of parameters) and for the BIC, it is $k \log(P)$.

From comparing the measures of fit given in Table 2, we may deduce two things. First, a diffusion or race model with equal item discrimination parameters fits much worse than the counterpart with free discrimination parameters. Since the model with equal discrimination parameters is nested within a model with unequal discrimination parameters, we can use a likelihood ratio test for comparing the fit of the two models. Both for the diffusion model ($L = 53.9, df = 9, p < .0001$) and for the regular race model ($L = 128.3, df = 9, p < .0001$), this leads to a rejection of the restricted model. Comparing the AIC and BIC result in an identical conclusion.

INSERT TABLE 2 ABOUT HERE

Second, the AIC and BIC criteria indicate that the race models are better fitting than the diffusion models. Moreover, there are two race models performing almost equally well: the delayed race model and the regular race model with random $\theta_p + \beta_i$. Based on the AIC there is no difference between the models, but the BIC seems to favor the delayed race model. Moreover, because the regular race model is nested within both the delayed race model (setting δ to zero in Equation (13) gives the regular race model) and the race model with random $\theta_p + \beta_i$, two likelihood ratio tests can be performed: one to test the hypothesis that $\delta = 0$ and the other to test the hypothesis that $\tau = 0$. The first test is significant ($L = 8.5, df = 1, p < .01$) leading to a rejection of the regular race model in favor of the delayed variant. The second test is also significant ($L = 4.3, df = 1, p < .02$;

note that in this case, we halved the p-value halved since we are testing a null hypothesis lying on the boundary of the parameter space, see Stram & Lee, 1994). Thus, the regular race model has to be rejected twice in favor of two other models (that do not have a nesting relation with respect to each other).

The parameter estimates for the three models with differing discrimination parameters (diffusion, race and delayed race model) and random $\theta_p + \beta_p$ can be found in Table 3. We have chosen to present the random $\theta_p + \beta_p$ versions of the model so that the information about the τ parameter is given in the table as well. When the random variability is removed from the models, the estimates of the remaining parameter estimates change only a little bit and their standard errors decrease slightly. The estimate for the constant residual RT parameter T_{er} is very similar across the three models (about 0.5s). The standard deviations of the population distributions for the θ_ps and ϵ_{pi}s differ a lot between the diffusion model and the two race models. That is because the two types of models measure the θ_ps on a different scale. Consequently, the β_is and α_is also differ in magnitude under the two models. If we evaluate the ratio η/τ, it is approximately 1.3 for the diffusion model but it is around 3 and 3.6 for the race and delayed race model. This indicates that there is more random $\theta_p + \beta_i$ in the diffusion model than there is in the race models.

Comparing the estimates of the parameters common to the regular and delayed race model reveals that they are very similar. The standard errors of the estimated α_is in delayed race model are a little bit larger than in the regular case, but the reverse holds for the estimated β_is and for the common parameters. As can be seen, the estimate of δ is negative, which means that endorsing a neuroticism item is slower than responding with the emotional stable alternative.

INSERT TABLE 3 ABOUT HERE

Model checking The delayed race model (without random $\theta_p + \beta_i$) seems to be the best fitting model, but that is only a relative conclusion: It does not show that the delayed race model actually fits the data and if that is the case, it is still possible that other models also fit the data relatively well. Therefore, the goodness-of-fit of each separate model to the data has to be assessed. As above with the testing of the 2PLM, we constructed test statistics tailored to evaluate specific aspects of the model and derived their reference distribution using parametric bootstrap. Four specific test statistics are defined below; two of the tests are summarized with a traditional p-value, the two other ones are entirely graphical.

As a first test, the overall goodness-of-fit for both models is evaluated with an omnibus test by comparing the observed deviance with the deviance derived from 100 model-based replicated data sets (see Equation (15)).

For the diffusion model without random drift, it turns out that the replicated deviance is often larger than the observed one ($\hat{p} = 0.91$). This is a somewhat odd observation and the explanation for it is that the observed data possess too few very large RTs than expected under the diffusion model. Consequently, this results in a lower deviance than expected because such extreme observations lower the likelihood. For the race model without random $\theta_p + \beta_i$, the deviances for the replicated data sets are in 58% of the cases larger than the observed deviance ($\hat{p} = 0.58$), indicating that the overall fit of the race model is very good. Also the delayed race model (without random $\theta_p + \beta_i$), with an estimated p-value is 0.47, fits the data very well. For the random drift diffusion model, the estimated p-value equals 0.99 (indicating an even worse fit than the variant without random drift). For the random version of the regular and delayed race model, the estimated p-values are 0.59 and 0.58, respectively.

Second, we looked at how well the three models can explain the conditional mean RTs displayed in Table 1. In Figure 3, the plots on the left-hand side shows the residuals (observed minus expected conditional mean RTs) versus the observed conditional mean RTs for the diffusion model and the random drift diffusion model, respectively. The points referring to an emotional stable answer are denoted by an asterisk and points referring to a neurotic answer by a circle; two points that belong to the same item are connected by a dotted line. The expected conditional mean RTs are estimated with model-based simulations. To assess the variability in the residual plot, we applied the same technique to four simulated data sets and these plots are shown at the right-hand side of the figure. Hence, the four plots on the right constitute the reference distribution for the residual plots. If the model fits, the observed and replicated plots should be indistinguishable.

For both diffusion models, there are two striking differences between the observed and replicated plots. First, in the observed plots, almost all residuals (there is only one small exception) corresponding to emotional stable responses are negative, meaning that their mean observed RTs are smaller than what is expected from the model. This is not the case in the replications, indicating a source of misfit. Second, the replicated conditional mean response times are from time rather large compared to the observed ones. For the observed plots, the values on the abscissa run roughly from 1.2s to 2.1s, while for the replicated data this range is often extends from 1.2s to almost 3s.

INSERT FIGURE 3 ABOUT HERE

In Panel A of Figure 4, residual plots are shown for the regular race model without (top row) and with random $\theta_p + \beta_i$ (bottom row). For the regular race model, the observed residual plot seems to contain larger residuals than the replicated versions. Moreover, the residuals for the neuroticism answers are all positive (with one small exception) while that is not the case in the replicated plots. The latter problem is less serious in the race model with random $\theta_p + \beta_i$, but the former (larger observed residuals) is still a problem. Note that the scale of the ordinate in Figure 4 is different from Figure 3 because the residuals from the race model are smaller than those from the diffusion model.

Panel B of Figure 4 contains the observed and simulated residual plots for the delayed race model without and with random $\theta_p + \beta_i$. As can be seen, the observed and replicated residual plots for the delayed race model without $\theta_p + \beta_i$ do not differ in a systematic way.

Note also that for the delayed race models (without random $\theta_p + \beta_i$), the expected conditional mean RTs is at most 0.20s off the mark (while for the regular race model with random $\theta_p + \beta_i$, the maximum observed residual is 0.23s). The three largest residuals in the delayed race model (without random $\theta_p + \beta_i$) fit are for the mean RTs for choosing *pessimistic* in item 6 (0.18s which gives a relative prediction error of $100\frac{0.19}{1.78} \approx 11\%$), *touchy* in item 7 ($-0.14$s resulting in a relative prediction error of about -6%) and *erratic* in item 10 (0.20s leading to a relative prediction error of about 12%).

INSERT FIGURE 4 ABOUT HERE

In a third test, we ignored the RT data and estimated the discrimination parameters only using the choice data (i.e., estimating the 2PLM). Then, the estimated discrimination parameters were correlated with the mean RTs (i.e., a correlation over items). For the observed data, this correlation was highly negative, -0.75. In a next step, the procedure was repeated for 100 model-based simulated data sets to see whether the models can reproduce the observed correlation. For the diffusion model, all replicated correlations between the discrimination parameter estimated from the choice data and the mean RTs were larger than the observed one and almost always positive, confirming the known fact that the diffusion model predicts a positive relation between the discrimination parameter and mean RT. The same was true for the diffusion model with random drift: It predicted systematically too high correlations. For the race model, 22% of the reproduced correlations were more negative than the observed one and for its random $\theta_p + \beta_i$ variant, this was 16%. For the delayed race model, 33% of the correlations based on the replicated data were smaller than the observed correlation and the random $\theta_p + \beta_i$ version had in 18% of the cases simulated correlations smaller than the observed one. Hence, we conclude that the correlation test shows a good fit for the race models because they are able to explain the negative correlation between item discrimination and mean RT, which is not the case for the diffusion model. The value of this test can hardly be overestimated because the strong negative correlation found between average RT and item discrimination as found in the data is totally irreconcilable with the diffusion model where a stronger discrimination implies a longer expected response time.

A final test of the models concerns the independence of choice and RT. For the diffusion and regular race model, the RT and choice response are independent conditional upon the true model parameter values. Because we do not know the true θ_ps, the sum score will be used as a proxy for θ_p instead. Again, the participants are divided in groups based on their sum score (see the section on the fit of the 2PLM). It can be assumed that each of these three groups are more or less homogeneous with respect to θ_p. Next, the difference between the mean RT for emotional stable responses (i.e., $X_{pi} = 0$) and neurotic responses (i.e., $X_{pi} = 1$) is computed for a particular item i in each group and then the differences are averaged over items.

The same procedure (constructing the three sum score groups, computing the average difference in conditional mean RTs within each group) is then applied to 100 simulated data sets under each of the six considered models (diffusion, regular race and delayed race model, all without and with random $\theta_p + \beta_i$). Next, the observed and replicated approximated CAFs are compared in six plots: one for each model with on the abscissa the three sum score groups and on the ordinate for each group the difference between the mean RTs for $X_{pi} = 0$ and $X_{pi} = 1$, averaged over the items. The results are shown in Figure 5; the two left panels contain the results for the diffusion models, the middle ones the results for the regular race models and the right panels for the delayed race models. The bold solid line represents the observed data, the bold dashed line the mean of the replicated data. The two thin solid lines are the 95% credibility interval bounds computed from the replicated data. (The thin dotted line is the zero line.)

For the observed data, the average time to endorse the neurotic alternative takes about 0.6s longer than to choose the emotional stable alternative in the Low group (people with a low position on the neuroticism continuum). On the other hand, in the High group (persons with a high position on the neuroticism continuum) the endorsement of the neurotic alternative is about 0.1s faster than choosing the emotional stable alternative. In the Middle group, we observe that endorsing the neurotic alternative takes about 0.2s longer compared to the emotional stable alternative.

As expected, the mean replicated CAF for the diffusion and regular race model without random $\theta_p + \beta_i$ almost coincides with the zero line. For the diffusion model without random drift, the observed CAF falls well within the 95% credibility bounds in all three sum score groups (mainly thanks to the very wide intervals), that is not the case for the race model. The mean replicated CAF of the delayed race model without random $\theta_p + \beta_i$ does not coincide with the zero line and it has a slightly steeper profile than the two previous competing models. The observed CAF falls nicely within the 95% credibility bounds for the delayed race model. For the random $\theta_p + \beta_i$ versions of the diffusion and regular race model, the mean replicated CAF clearly looks more similar to the observed one and for the regular race model, the observed CAF also falls within the 95% confidence bounds. For the delayed race model with random $\theta_p + \beta_i$, the situation is very similar to graph on the top row. All race models produce narrower confidence bands than the diffusion models.

INSERT FIGURE 5 ABOUT HERE

Interpretation The presented model evaluation rules out the diffusion model as a viable underlying response model. The delayed race model without random $\theta_p + \beta_i$ is retained as the best fitting model (lowest AIC and BIC, and no deviations with respect to the model diagnostics) and we will focus in the interpretation on this model.

Research on the representation of personality traits in memory (Klein & Loftus, 1993; Klein, Loftus, Trafton, & Fuhrman, 1992) has revealed that traits are often linked to exemplar information (concrete behaviors) and/or abstract information (the behavioral information is already summarized in a few abstract representations linked to the self-descriptive trait). If the response process depends (even only partially) on the use of exemplar information, then one expects that correlations can be found between the parameter estimates and measures related to the behavioral information. If these correlations are found, they will aide in the interpretation of the race model parameters. Recall that the discrimination parameter in the race model is directly function of the dependency between the information nodes in the decision process. From the estimated discrimination indices in Table 3, we can derive that, for instance, more independent information is available for the trait pair *panicky* vs calm ($\hat{\alpha} = 0.71$) than for the pair *touchy* vs imperturbable ($\hat{\alpha} = 0.14$).

Now is there a compelling reason why some trait pairs discriminate more than others? Tuerlinckx and De Boeck (2004) argued that a variable that might help to explain these differences in discrimination is category breadth. Category breadth has many aspects (Buss & Craik, 1983) and two are relevant for this research. First, it refers to the category volume and as such it is defined as the number of distinct behaviors subsumed by a trait (Hampson, John, & Goldberg, 1986). A second aspect of category breadth refers to the amount of dependency among the behaviors belonging to a trait or the number of distinct behaviors subsumed by the trait (Hampson et al., 1986). If the response process is based on behavioral exemplars, then the second aspect indicates that category breadth should correlate with the discrimination parameter from the race model, because the discrimination parameter is inversely related to the dependency between the information nodes. Category breadth is usually defined for single traits, but it can be extended easily to pairs of traits. We have collected ratings for the average breadth of the trait pairs in our example here (for details of the rating procedure, see Tuerlinckx & De Boeck, 2004). The category breadth ratings and the estimated discrimination parameters correlate significantly ($r = 0.56, df = 8, p < .05$, one-sided), supporting our interpretation of the discrimination parameter.

The correlation between category breadth and the discrimination parameter is not perfect of course. First of all, only 10 items were considered in this research (but with more than 40 items, the correlation is still around 0.6; see Tuerlinckx & De Boeck, 2004). Second, it is likely that the decision process is only partially based on exemplar information. Other information that is likely to be used in the decision process are previously made trait inferences. Third, the other relevant aspect of category breadth (the category volume) plays a role too. In the two race models, the number of information nodes considered for a particular item by a person is equal to $N_A + N_B$ (see

Equation (5)). It is clear that for a given θ, the number of nodes increases if the absolute value of β increases. Therefore, we expect $|\beta|$ to be related to category breadth too, because $|\beta|$ is reflects the category volume. Also this correlation turns out to be significant ($r = 0.63, df = 8, p < .05$, one-sided). Moreover, if we reconsider the correlation between the discrimination parameter and category breadth, but now controlling for $|\beta|$, the latter correlation becomes even stronger ($r = 0.75, df = 7, p < .01$, one-sided). These results show that the race model interpretation of the discrimination parameter is supported by substantive evidence from personality psychology.

Discussion

In this paper we derived the 2PLM as the choice probability part of two different kinds of stochastic models: an unbiased diffusion model and two race models with dependencies between the racing information (one with independence between choice and RT and the other without the latter assumption). We obtained two novel interpretations of the discrimination parameter. In the case of the diffusion process, the discrimination parameter is a quantification of the amount of information that has to be collected before a response to an item is given. In the case of the race models, the discrimination parameter is a function of the dependence among the bits of information on which the decision is based. The two types of stochastic models provide, besides the choice probabilities, also a RT distribution which makes it possible to distinguish empirically between them.

It is also shown that one might add random variability to the sum of the person and item parameters, $\theta_p + \beta_i$, and that this still leads to a model for the choice data that is practically indistinguishable from the 2PLM. Adding random variability to $\theta_p + \beta_i$ leads to a model with person-by-item interactions. With binary data only, such an interaction model is overparametrized, but when additionally RTs are collected, the contribution of the interaction may be tested. However, the estimate of the interaction variance (τ) may hinge on the particular model assumptions, as was the case in our example. Therefore, model checking should proceed the interpretation of the model.

For the particular data set analyzed in this paper, it was demonstrated that the delayed race model has a better fit than the diffusion and the regular race model (even adding random error to $\theta_p + \beta_i$ does not lead to a better fit for the latter two). Moreover, it appears that the delayed race model fits most aspects of the data quite well, while the diffusion and regular race model fail to do so for important aspects of the data. In addition, the discrimination parameter under the race model can be interpreted and this interpretation is validated by external information.

Of course, the fact that the delayed race model is more appropriate for the data considered in this paper does not imply at all that the (delayed) race model will always outperform the diffusion model. The diffusion model is shown to be a plausible response model in many psychological tasks. One of the reasons we find that a race model fits better, may be related to the fact that the mean RTs on the items run into several seconds, while in the more traditional applications of diffusion models the average RTs are below one second (Ratcliff, 2002). It is possible that the diffusion model is not very well suited for fitting data from tasks that produce such a wide range of RTs.

The diffusion model presented in this paper is a version of a model that has been applied frequently in two-alternative forced choice tasks (see e.g., Ratcliff, 1978; Ratcliff et al., 1999). However, the presented version of the race model is rather new and an important concern regarding these race models are the results of Marley and Colonius (1992) and Dzhafarov (1993). They have shown that any set of choice probabilities and RTs can be modeled by an arbitrary race model. Consequently, this means that the specific setup of a race model should be motivated by external factors (other than the data). Our specific formulation of the race model is chosen for two reasons. First, the choice probability under the race model was constrained to be the 2PLM. This constraint already rules out several other possible race models. Second, the conceptualization we choose is especially suited for the example discussed. From the research on the representation of personality traits (Klein et al., 1992; Klein & Loftus, 1993; see above), it seems plausible to consider that the information nodes in the race model exactly code for these discrete bits of information representing exemplar (but also abstract) information concerning a personality trait.

The interpretations for the discrimination parameter we propose in this paper is especially suited for two-alternative forced choice items for which the response alternatives are two concrete responses (and not a class of possible responses). This requirement excludes ability items. Although an ability item could be conceived as a two-alternative forced choice item where the person has to choose between a correct response and a wrong response, there are often many wrong responses and only one correct response, so that the "wrong" response is actually a heterogeneous class of responses.

As already mentioned above, the connection outlined between the 2PLM and the two stochastic models may have implications for the 2PLM itself, such that in some situations there may be a need to alter the 2PLM. For example, consider the failure by the diffusion and regular race model to account for the dependency between choice and RT. As a consequence, there is a need to expand the models and we have considered such possibilities (adding random error to $\theta_p + \beta_i$ and introducing the delayed race model). However, there are many other ways one could adjust the models such that they become more realistic. We did not pursue other courses of action here because such changes of the joint distribution of RTs and choice probabilities may produce models whose marginal choice probability component deviates strongly from the 2PLM. The main objective of our research was to construct models that may help in the process of understanding the meaning of the discrimination parameter, rather than proposing alternatives for the 2PLM. Moreover, we have shown that the 2PLM fits the choice data, hence there was no urge to modify the choice model. However, in other applications it may prove necessary to reject the 2PLM on empirical grounds as a model for the choice probabilities in order to arrive at a well-fitting joint model for choices and RTs.

In conclusion, we believe it is worth looking for good and testable interpretations of the discrimination parameter. Without such an interpretation, the 2PLM is merely reduced to a curve fitting tool, with the item discrimination as a parameter that is inserted to give the model more flexibility but without a theoretical and substantive basis for the discriminations to differ from one another.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Proceedings of the symposium on applications of statistics* (pp. 27–47). Amsterdam, The Netherlands: North Holland.

Birnbaum, A. (1968). Some latent trait models. In F. M. Ford & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–424). Reading: Addison-Wesley.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review*, *90*, 105–126.

Colonius, H. (1981). A new interpretation of stochastic test models. *Psychometrika*, *46*, 223–225.

Cox, D. R., & Miller, H. D. (1970). *The theory of stochastic processes.* London: Methuen.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, *41*, 260–274.

Dzhafarov, E. N. (1993). Grice-representability of response time distributions. *Psychometrika*, *58*, 281–314.

Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement.* Englewood Cliffs: Prentice Hall.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika*, *64*, 407–433.

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models: Foundations, recent developments, and applications.* New York: Springer-Verlag.

Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review*, *75*, 359–373.

Hampson, S. E., John, O. P., & Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: Studies of asymmetry in judgments of trait implications. *Journal of Personality and Social Psychology*, *51*, 37–54.

Hougaard, P. (1986a). Survival methods for heterogeneous populations derived from stable distributions. *Biometrika*, *73*, 387–396.

Hougaard, P. (1986b). A class of multivariate failure time distributions. *Biometrika*, *73*, 671–678.

Joe, H. (1997). *Multivariate models and dependence concepts.* London: Chapman & Hall.

Klein, S. B., & Loftus, J. (1993). The mental representation of trait and autobiographical knowledge about the self. In T. K. Srull & J. R. S. Wyer (Eds.), *The mental representation of trait and autobiographical representation of the self (advances in social cognition, Volume V)* (pp. 1–49). Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, S. B., Loftus, J., Trafton, J. G., & Fuhrman, R. W. (1992). Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology*, *63*, 739–753.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison Wesley.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Marley, A. A. J., & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology*, *36*, 1–20.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Naylor, J. C., & Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, *31*, 214–225.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–104.

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, *92*, 212–225.

Ratcliff, R. (2002). Diffusion and random walk models. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 3668–3673). Amsterdam: Elsevier.

Ratcliff, R., & Rouder, J. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*, 261–300.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

SAS Institute Inc. (1999). *SAS System V8 for Windows* [Computer program]. Carey, NC: SAS Institute Inc.

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects mode. *Biometrics*, *50*, 1171–1177.

Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behavior Research Methods, Instruments, & Computers*, *in press*.

Tuerlinckx, F., & De Boeck, P. (2004). Stochastic models for personality trait judgments. *Unpublished manuscript, University of Leuven*.

Author Note

Address correspondence to Francis Tuerlinckx, Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium. E-mail: francis.tuerlinckx@psy.kuleuven.ac.be. The author thanks Frank Rijmen, the Editor, the Associate Editor and two reviewers for useful suggestions. The first author is a postdoctoral researcher for the Fund of Scientific Research - Flanders. Preparation of this article was supported in part by the BOF grant GOA/2000/02 from the K.U.Leuven, Belgium and by the NSF grant SES-0084368.

Footnotes

[1]The random $\theta + \beta$ extension will not be denoted as a random drift race model because there is no drift rate in the race model.

[2]Upon request the SAS code can be obtained from the first author.

Table 1
Descriptive statistics (proportion neurotic alternatives endorsed, conditional mean RTs and standard deviations) for the 10 items.

	Item		Proportion neurotic choices (SD)	Mean RT Neurotic (SD)	Mean RT Emot. stable (SD)	Sign. Diff.[a]
	Neurotic alternative	Emotional stable alternative				
1	*nervous*	relaxed	0.35	1.52 (0.62)	1.30 (0.39)	*
2	*panicky*	calm	0.29	1.45 (0.57)	1.26 (0.47)	
3	*despondent*	lively	0.13	1.80 (0.89)	1.45 (0.64)	
4	*depressed*	optimistic	0.14	1.57 (0.91)	1.17 (0.55)	*
5	*timid*	intrepid	0.49	1.88 (0.99)	1.65 (0.62)	
6	*pessimistic*	optimistic	0.15	1.78 (0.99)	1.20 (0.59)	*
7	*touchy*	imperturbable	0.58	2.05 (0.90)	2.04 (1.15)	
8	*uncontrolled*	controlled	0.14	1.80 (0.63)	1.51 (0.68)	
9	*melancholy*	light-footed	0.57	1.73 (0.58)	1.66 (0.64)	
10	*erratic*	balanced	0.27	1.72 (1.04)	1.24 (0.47)	*

Note. [a] An asterisk in this column indicates a significant difference ($p < 0.05$) between the logarithm of the two conditional mean RTs for the item using a two-samples t-test.

Table 2
Fit statistics for various models.

		-2ℓ	AIC	BIC
1	Diffusion model with unequal item discriminations	2420	2464	2521
2	Diffusion model with equal item discriminations	2474	2500	2534
3	Diffusion model with random drift	2405	2451	2510
4	Race model with unequal item discriminations	2320	2369	2430
5	Race model with equal item discriminations	2449	2479	2517
6	Delayed race model	2312	2362	2426
7	Race model with random $\theta_p + \beta_i$	2316	2362	2428
8	Delayed race model with random $\theta_p + \beta_i$	2310	2366	2430

Table 3
Parameter estimates (and asymptotic standard errors below) for the diffusion model, regular race model and delayed race model with $\theta_p + \beta_i$.

		Diffusion		Race		Delayed race	
		General parameters					
	T_{er}	0.48		0.51		0.51	
		(0.01)		(0.003)		(0.003)	
	η	0.80		4.66		4.26	
		(0.12)		(1.65)		(1.60)	
	τ	0.58		1.46		1.17	
		(0.10)		(0.63)		(0.58)	
	γ	–		1.90		1.89	
				(0.06)		(0.06)	
	λ	–		0.14		0.15	
				(0.03)		(0.03)	
	δ	–		–		−0.21	
						(0.08)	
Item		Item-specific parameters					
		α	β	α	β	α	β
nervous	relaxed	2.41	−0.47	0.56	−1.89	0.61	−1.59
		(0.14)	(0.16)	(0.19)	(0.89)	(0.22)	(0.81)
panicky	calm	2.03	−0.63	0.67	−2.11	0.71	−1.81
		(0.10)	(0.16)	(0.21)	(0.87)	(0.23)	(0.80)
despondent	lively	2.93	−1.31	0.41	−5.86	0.43	−5.29
		(0.19)	(0.18)	(0.15)	(2.17)	(0.16)	(2.05)
depressed	optimistic	2.48	−1.48	0.60	−4.87	0.64	−4.29
		(0.15)	(0.20)	(0.19)	(1.79)	(0.21)	(1.64)
timid	intrepid	2.73	0.02	0.36	0.38	0.40	0.42
		(0.15)	(0.15)	(0.15)	(0.66)	(0.17)	(0.60)
pessimistic	optimistic	2.54	−1.41	0.57	−4.92	0.61	−4.35
		(0.16)	(0.20)	(0.18)	(1.80)	(0.20)	(1.65)
touchy	imperturbable	2.94	0.19	0.11	3.15	0.14	2.57
		(0.16)	(0.14)	(0.10)	(3.12)	(0.12)	(2.51)
uncontrolled	controlled	3.12	−1.28	0.37	−6.36	0.38	−5.98
		(0.21)	(0.18)	(0.14)	(2.41)	(0.15)	(2.38)
melancholy	light-footed	2.81	0.19	0.36	1.21	0.41	1.20
		(0.17)	(0.15)	(0.15)	(0.85)	(0.18)	(0.79)
erratic	balanced	2.38	−0.80	0.61	−2.84	0.65	−2.45
		(0.13)	(0.17)	(0.19)	(1.06)	(0.21)	(0.97)

Figure Captions

Figure 1. A schematic representation of the race model. The circles at the top are the information nodes racing to activate the two decision nodes (at the bottom), which are corresponding to alternatives A and B. The horizontal lines connecting the information nodes refer to possible dependencies between the information nodes (with dependency parameter ω).

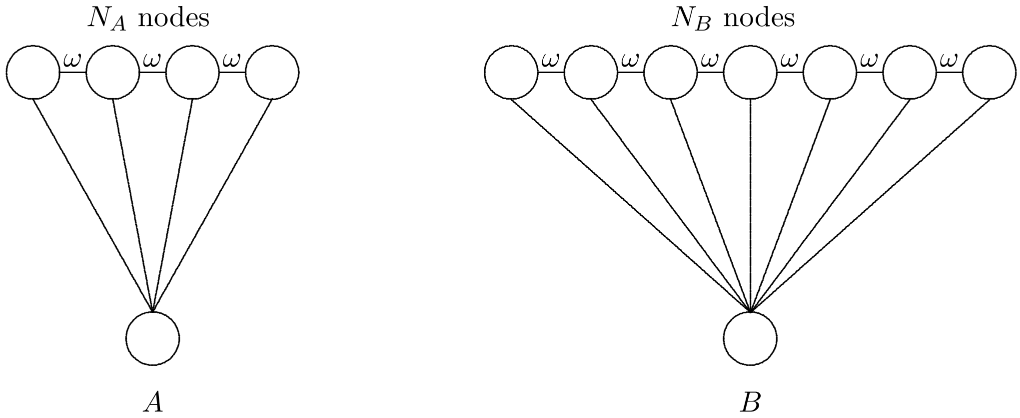
Figure 2. Graphical goodness-of-fit test for the 2PLM. For each of the three sum score groups (Low with $s_p = 0, 1$, Medium with $s_p = 2, 3, 4$, and High with $s_p = 5, 6, 7, 8, 9, 10$ where s_p is the sum score) the proportion endorsed neurotic responses for each item (separate panels) in the observed data were computed (denoted by the crosses). The same was done for 100 simulated data sets and 95% confidence intervals were computed for the replicated proportions (the thick lines).

Figure 3. The two most left-hand side plots contain the residual plots for the diffusion model (top row) and for the diffusion model with random drift (bottom row) as a function of the observed conditional mean RTs. Asterisk points refer to the mean response times and residuals of the emotional stable responses and circles to those of the neurotic responses. The reference distribution of the residuals is simulated by generating under each model four replicated data sets which have been subjected to the same kind of analysis as the observed data. These replicated residual plots are for each model shown the four right-hand side plots.

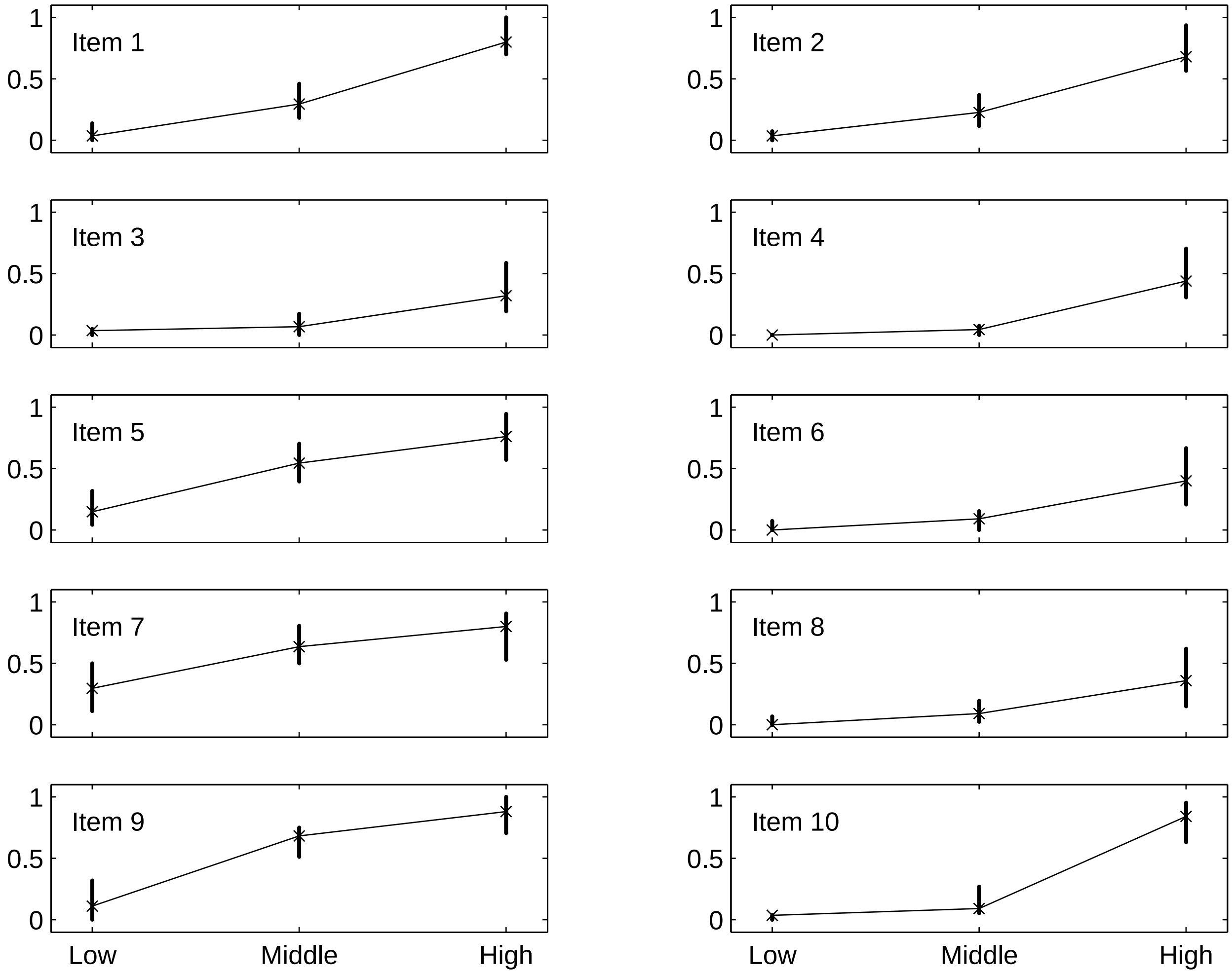
Figure 4. Panel A: Observed (left-hand side plot) and replicated (four right-hand side plots) residual plots for the regular race model (top row) and race model with random $\theta + \beta$ (bottom row) as a function of the conditional mean RTs. Panel B: Observed (left-hand side plot) and replicated (four right-hand side plots) residual plots for the

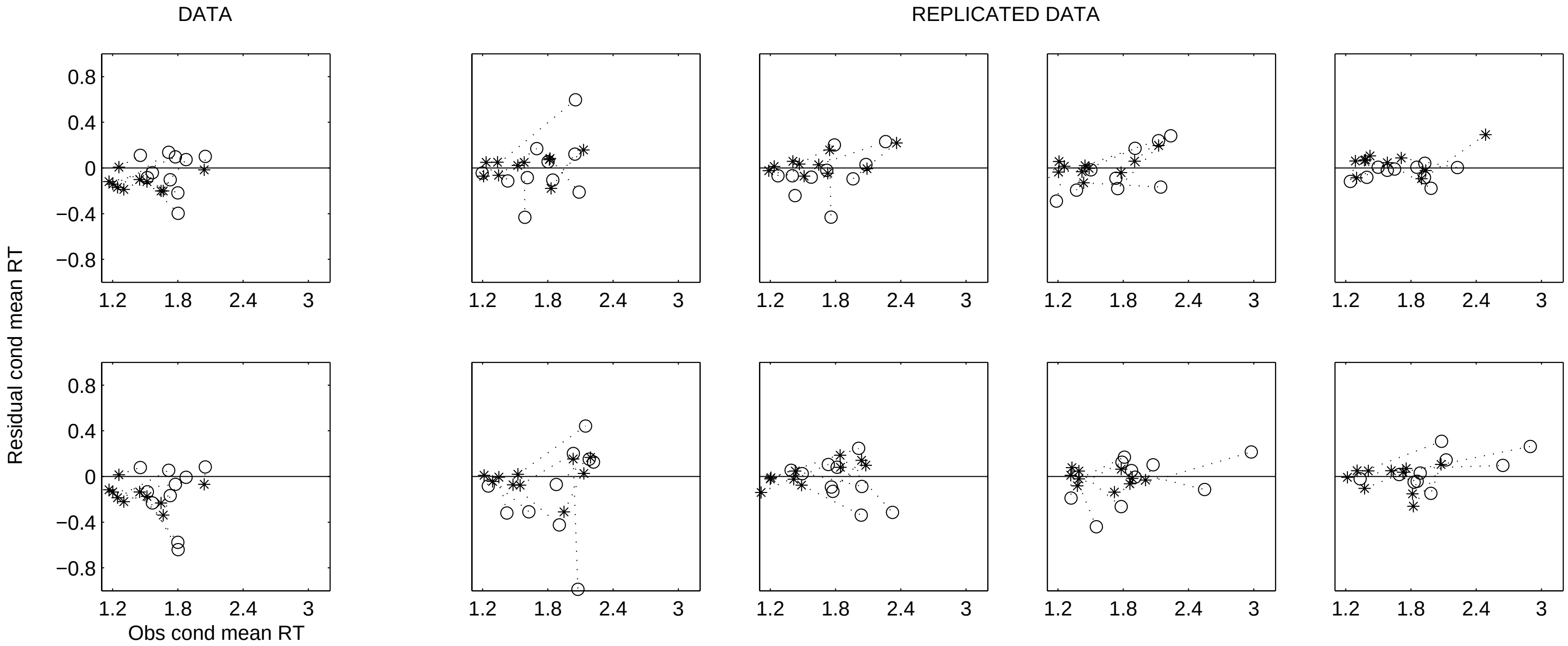
delayed race model (top row) and delayed race model with random $\theta + \beta$ (bottom row) as a function of the conditional mean RTs. In both panels, asterisk points refer to emotional stable responses and circles to neurotic responses (see also caption of Figure 3).

Figure 5. For each of the three sum score groups (Low, Medium, and High; see also the caption of Figure 2 for a definition of the groups), the difference in mean RT for choosing the emotional stable alternative and the mean RT for choosing the neurotic alternative is plotted. The bold solid line with the stars represents the observed relation, the outer two thin solid lines refer to the 95% credibility interval bounds (based on 100 simulations). The bold dashed line with the crosses represents the average profile of the simulated data. Panel (a) contains the results of the diffusion model; panel (b) of the regular race model, (c) of the delayed race model, (d) of the diffusion model with random drift, (e) of the regular race model with random $\theta_p + \beta_i$, and (f) of the delayed race model with random $\theta_p + \beta_i$.

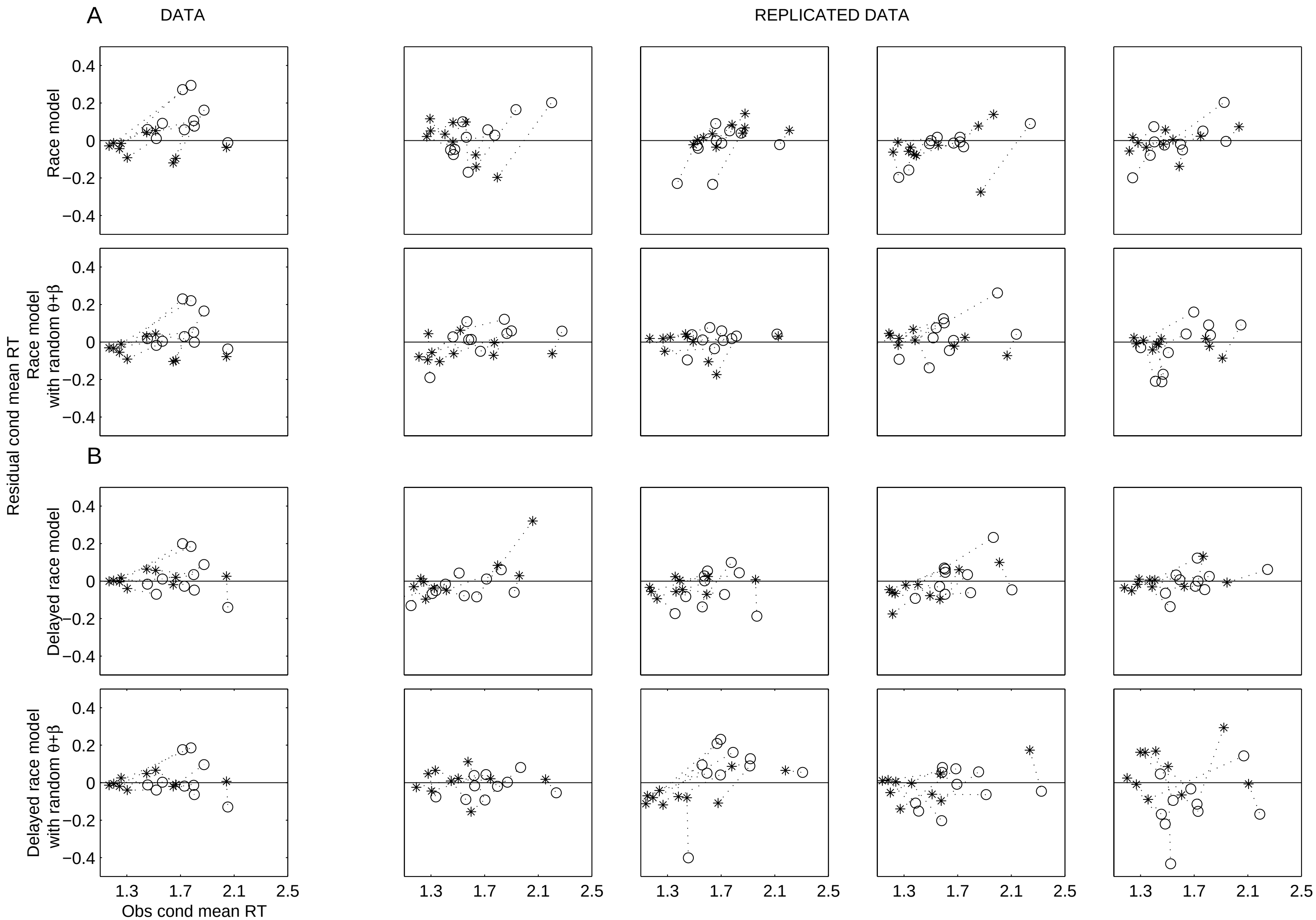


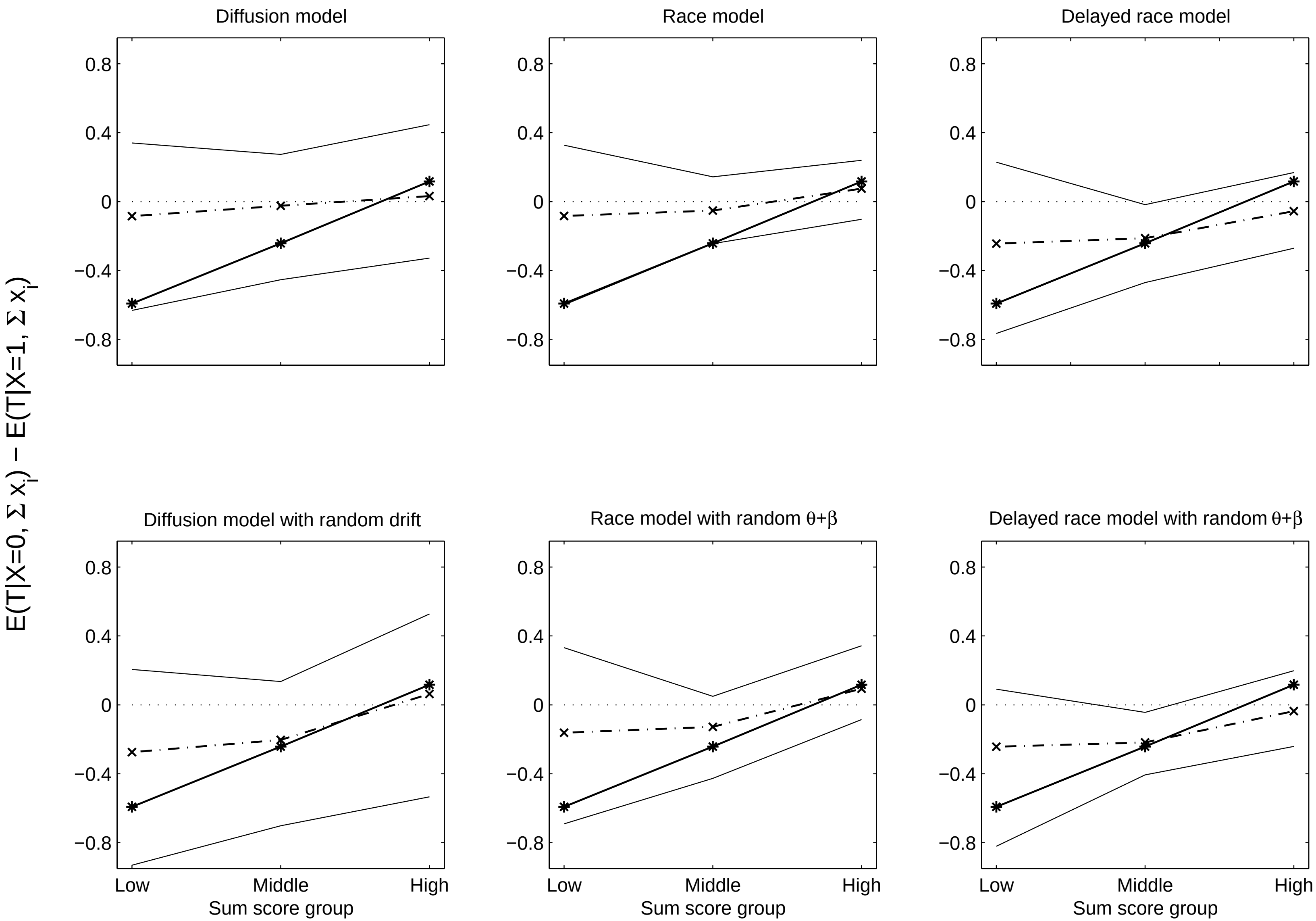
N_A nodes
ω
ω
ω
A
N_B nodes
ω
ω
ω
ω
ω
ω
B


Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7
Item 8
Item 9
Item 10
1
0.5
0
Low
Middle
High



DATA
REPLICATED DATA
Residual cond mean RT
Obs cond mean RT
0.8
0.4
0
−0.4
−0.8
1.2
1.8
2.4
3


A
DATA
REPLICATED DATA
Race model
Race model
with random θ+β
Residual cond mean RT
B
Delayed race model
Delayed race model
with random θ+β
0.4
0.2
0
−0.2
−0.4
1.3
1.7
2.1
2.5
Obs cond mean RT


E(T|X=0, Σ x_i) − E(T|X=1, Σ x_i)
Diffusion model
Race model
Delayed race model
Diffusion model with random drift
Race model with random θ+β
Delayed race model with random θ+β
0.8
0.4
0
−0.4
−0.8
Low
Middle
High
Sum score group