Chapter 1

Rasch model, log-normal model, and their specification for analyzing IAT data

This Chapter is organized in two main sections.

In the first section, the Rasch model and the log-normal model are briefly outlined.

Then, the similarities between Rasch model and Generalized Linear (Mixed Effects) model are described, as well as the rationale according to which it is possible to estimate Rasch model parameters by employing Generalized Linear Mixed Models (GLMMs) with a *logit* link function from accuracy responses. A similar logic is applied for the estimation of the log-normal model parameters from the log-transformed time responses by using Linear Mixed Effects Models (LMMs).

In the second section, the random structures of the GLMMs and the LMMs used for estimating Rasch model and log-normal estimates from IAT accuracy and log-time data, respectively, are presented.

Three random structures for accuracy responses (Rasch model), as well as three random structures for log-time responses (log-normal model) are introduced. The first one is the simplest one, and it is considered the Null model against which the models with the other two random structures are compared. The second and third models have the same level of complexity. They differentiate each other according to the random factor on which the multidimensionality is allowed, either the respondents (Model 2) or the stimuli (Model 3). Therefore, the best fitting model and the estimation of the Rasch model and log-normal model estimates depend on the variability of the observed data.

For illustration purposes, the Rasch model is initially presented with the typical notation for its parameters, namely β indicating persons' abilities and δ indicating items difficulty. However, since also the item parameter of the log-normal model is indicated with δ, a different notation for the Rasch model, more similar to the one used in Item Response Theory in general, is employed. Consequently, respondents' parameters are indicated with Greek letter θ and item parameters with the Latin letter b. Respondents are indicated with the subscript p ($p = \{1, \ldots, P\}$) and stimuli/items with the subscript s ($s = \{1, \ldots, S\}$). In the specification of Linear Mixed Effects Models, the single observation on each respondent p on each stimulus s in each associative condition c ($c = \{1, \ldots, C\}$) is identified by i ($i = \{1, \ldots, I\}$).

1.1 Modeling dichotomous responses

According to Item Response Theory (IRT) models, the observed response to an item can be explained by a common characteristic shared by both the person and the item, which lies on the same latent trait (?, ?). IRT scoring accounts for the moderation of item characteristics in explaining the relationship between the person's latent trait, often identified with θ, and the observed response. IRT models can be distinguished according to the number of parameters used for describing the item characteristics (e.g., ?, ?).

The simplest one is the 1-Parameter logistic model (1PL, Equation 1.1), which shares some characteristics with the Rasch model (?, ?). According to this model, the probability of a correct response is a function of the respondent's characteristic θ and an item characteristic, defined as difficulty, b:

$$P(x_{ps} = 1|\theta_p, b_s) = \frac{exp[a(\theta_p - b_s)]}{1 + exp[a(\theta_p - b_s)]} \tag{1.1}$$

Difficulty b can be defined as the amount of latent trait θ that is needed for having a higher probability of choosing the correct response over the incorrect response. The discriminatory power of the item (i.e., parameter a, the ability of the item to distinguish between respondents with a high level of θ from those with a low level of θ) does not have a subscript indicating the specific items because it is equal for all the items (?, ?). Besides the difference in the notation of the parameters (θ_p and b_s in the 1PL, β_p and δ_s in the Rasch model, for identifying the respondents' ability and item difficulty, respectively), 1PL model and Rasch model are mathematically equivalent.

The 2PL model (Equation 1.2) (?, ?) also considers the influence of each item discrimination power (parameter a) in explaining the relationship between the respondent's ability and the observed response:

$$P(x_{ps} = 1|\theta_p, b_s, a_s) = \frac{exp[a_s(\theta_p - b_s)]}{1 + exp[a_s(\theta_p - b_s)]} \tag{1.2}$$

As it can be seen from Equation 1.2, parameter a changes the relationship between respondent's parameter θ and the item difficulty parameter b. The larger the value of a, the lower the overlap between the distributions of the response variables of two respondents with different values of θ. In this sense, parameter a can be interpreted as the discriminating power of the item. Item with large value of a are better able to discriminate respondents with different levels of θ.

Both 1PL and 2PL models assume a lower asymptote of 0 (i.e., the value taken by the function as θ approaches $-\infty$, defining the probability of an incorrect response) and an upper asymptote of 1 (i.e., the value taken by the function as θ approaches $+\infty$, defining the probability of an correct response). The Rasch model assumes a lower asymptote at 0 as well. The assumption of the lower asymptote approaching zero implies that respondents with extremely low levels of ability have an extremely low probability of endorsing the correct response. Conversely, assuming an upper asymptote of 1 assumes implies that respondents with extremely high ability level have an extremely high probability of endorsing the correct responses.

However, there might be cases in which respondents with an extremely low level of ability endorse the right response just out of luck (lucky guess), or that respondents with an extremely high level of ability endorse the incorrect response just out of distraction (careless error). In the first case, the lower asymptote cannot approach zero anymore, since even respondents with a low level of ability have a slight probability of endorsing the correct response. In the latter one, the upper asymptote has to be moved downward because even respondents with an extremely high level of ability does not the probability of correctly respond to each item.

The 3PL and 4PL models have been introduced for modeling these occurrences, respectively.

The 3PL model (Equation 1.3) (?, ?) adds a third parameter (c) to explain the relationship between θ and the observed response:

$$P(x_{ps} = 1|\theta_p, b_s, a_s, c_s) = c_s + (1 - c_s)\frac{exp[a_s(\theta_p - b_s)]}{1 + exp[a_s(\theta_p - b_s)]}, \tag{1.3}$$

where c models the probability that a respondent with a low level of ability guesses the correct response. Parameter c hence moves upward the lower asymptote.

The 4PL model (Equation 1.4) (?, ?) adds a fourth item parameter (e) for moderating the relationship between the item and the respondent:

$$P(x_{ps} = 1|\theta_p, b_s, a_s, c_s, e_s) = c_s + (e_s - c_s)\frac{exp[a_s(\theta_p - b_s)]}{1 + exp[a_s(\theta_p - b_s)]}, \tag{1.4}$$

where e is the parameter that describes the likelihood that respondents with an ability level higher than the item difficulty would endorse the incorrect response (i.e., careless error). As it can be seen from $(e_s - c_s)$ of Equation 1.4, the upper asympote is no more defined by the probability of giving the correct response (1), but by parameter e_s.

1.1.1 The Rasch model

The Rasch model (?, ?) is not a special case of the 2PL model, even though it is mathematically equivalent to the 1PL model (?, ?). Rasch model drops completely parameter a, and it is specified in terms of *log-odds* (i.e., *logits*, the natural logarithm of the odds).

Despite the 1PL IRT model and the Rasch model are mathematically equivalent, the notational system used for their parameters is different. In the Rasch model, the item parameter is described by the Greek letter δ and the person's parameter is described by β. In this section, the typical notation of the Rasch model is used. However, in Section 1.4 the notation typical of IRT models is used to distinguish the Rasch model parameter estimates from the estimates of the log-normal model and those of the GLMMs.

The starting point for the development of the dichotomous Rasch model (?, ?) involves the engagement of a person p on an item s to produce a response x_{ps} (?, ?). The engagement results from a single variable that is a common property shared by both the person and the item. The item variable is supposed to trigger the same person variable in all respondents. For instance, for assessing mathematics proficiency the items must contain some degree of mathematics ability. To give the correct response, persons must engage with the mathematics proficiency required by the item.

The engagement between persons and items results in the observed responses x_{ps}, which can be represented in a p ($p = \{1, \ldots, P\}$, persons) $\times$ items s ($s = \{1, \ldots, S\}$, items) response matrix $\boldsymbol{X}$ (Table 1.1).

Table 1.1: Response matrix $p \times s$, starting point for estimating the Rasch model.

		Items						
		1	2	...	k	...	s	
	1	x_{11}	x_{12}	...	x_{1k}	...	x_{1k}	r_1
	2	x_{21}	x_{22}	...	x_{2k}	...	x_{2k}	r_2
	$\vdots$	...	...	...	...	...	...	$\vdots$
Persons	v	x_{v1}	x_{v2}	...	x_{vk}	...	x_{vk}	r_v
	$\vdots$	...	...	...	...	...	...	$\vdots$
	p	x_{p1}	x_{p2}	...	x_{pk}	...	x_{ps}	r_p
		s_1	s_2	...	s_k	...	s_s	

Each cell represents the response of person p to item s. The response is a dichotomous response that can only the values $x_{ps} = 0$ (incorrect response) or $x_{ps} = 1$ (correct response).

The across-columns sum r_p (i.e., number-correct) represents the total score of each respondent (i.e., the total number of correct responses given by the respondent), regardless of the specific pattern with which the correct responses were given. The number-correct is a sufficient statistic for estimating the person's parameter β (?, ?, ?). Two respondents might have the same number-correct obtained with different patterns of correct responses. Since the specific pattern does not matter for the determination of the number-correct, the two respondents with same number-correct obtained with different pattern will have the same person estimate β. This feature is one of the peculiarities that distinguish the Rasch model from other IRT models. For instance, in the 2PL two respondents with the same number-correct might not have the same level of θ because the relationship between the item and the respondent's estimate is moderated by the discrimination parameter a.

Similarly, the across-rows sum s_s (i.e., proportion-correct) represents the total score of each item (i.e., number of correct responses obtained by each stimulus), regardless of the specific pattern with which the responses were given. The proportion-correct is a sufficient statistic for estimating the item difficulty parameter δ. Two items might have the same proportion-correct resulting from different pattern of responses, but since the specific pattern is not relevant for the determination of the proportion correct, the two items will have the same item estimate δ.

The concept of sufficient statistics is directly related to that of Local independence, which is further illustrated in paragraph 1.1.1.

The observed response in each cell x_{ps} hence depends on both persons' characteristics and items characteristics. Characteristics of both persons and stimuli can be located on a specific point of the latent trait, which is the common variable they share and that is assessed. The locations of each respondent p on the latent trait are described by parameter β_p. The location of each stimulus on the latent trait is described by parameter δ_s. While the observed response for each combination of $p \times s$ can take only the value 0 and 1, the parameters β_p and δ_s can take any real values from $-\infty$ to $+\infty$.

The fact that persons and items are located on the same latent traits is the great advantage of the Rasch model. By sharing the same latent trait, it is possible to directly compare persons' estimates with items estimates, and hence a measure of the distance between them can be obtained. Therefore, it is possible to predict the probability that a person with a certain level of β has of correctly respond to an item with a certain level of δ. Therefore, since the observed response is a function of respondents' and stimuli characteristics located on the same latent trait, it is possible to speculate that a respondent would correctly respond to stimuli below his/her level of ability β_p (i.e., the probability of a correct response is higher than 50%):

$$\text{If } (\beta_p - \delta_s) > 0 \text{ then } P(x_{ps} = 1) > 0.50. \tag{1.5}$$

Consequently, when the location of the item is above the location of the person, the probability that a correct response is given is below 50%:

$$\text{If } (\beta_p - \delta_s) < 0 \text{ then } P(x_{ps} = 1) < 0.50. \tag{1.6}$$

Assuming that respondents and items lies on the same latent variable also implies that the probability of a correct response can be related with the difference between respondents parameters and item parameters, as shown in Equation 1.5 and Equation 1.6.

However, the probability of a correct response is bounded between 0 and 1, while the parameters, and hence their difference, can vary between $-\infty$ and $+\infty$. The difference between respondents' and items parameters can be forced to only positive numbers by using the exponential distribution:

$$0 \leq exp(\beta_p - \delta_s) < +\infty \tag{1.7}$$

Still, the difference between person's ability and item difficulty cannot be used to predict a probability, because it is allowed to take any positive value from 0 to $+\infty$. To map the difference between respondents and item parameters on the same scale of the probability, and hence to use them for predicting the probability of a response given respondents' ability and item difficulty, Equation 1.7 can be standardized by $1 + exp(\beta_p - \delta_s)$. The probability for a correct response for a given β_p and a given δ_p can hence be expressed as:

$$P(x_{ps} = 1|\beta_p, \delta_s) = \frac{exp(\beta_p - \delta_s)}{1 + exp(\beta_p - \delta_s)}, \tag{1.8}$$

which is the typical formulation of the Rasch model for the probability of a correct response.

As said before, the Rasch model was originally formulated in terms of odds and *log-odds*. Equation 1.8 can hence be rewritten in terms of *log-odds*:

$$\beta_p - \delta_s = \ln\left(\frac{P(x = 1|\beta_p, \delta_s)}{1 - P(x = 1|\beta_p, \delta_s)}\right), \tag{1.9}$$

or, by applying the properties of logarithms to Equation 1.8, the Rasch model can be re written:

$$P(x_{ps} = 1|\beta_p, \delta_s) = \frac{1}{exp(\delta_s - \theta_p)} \tag{1.10}$$

The formulation in Equation 1.8 makes clear that the only thing that matters for the estimation of the expected probabilities is the difference between β_p and δ_s. This difference expresses the distance between the location of respondent p from the location of stimulus s on the latent trait. The probability of a correct (incorrect) response changes according to the distance between respondent's and item locations. The probability of a correct response is 50$\%$ when respondent's location is equal to item location. The variance for the expected probabilities of responses when respondent's and stimuli locations correspond is maximized. The more (less) the respondent's location is above the item location, the higher (lower) the probability of a correct response (see Equation 1.5). Also the opposite holds true. The more (less) the location of the respondent is below the item location, the higher (lower) the probability of an incorrect response (see Equation 1.6). The relationship between the respondent's parameter and the item parameter defines the cumulative nature of the Rasch model.

The Rasch model assumes a logistic probability function, and the measurement units of the respondents' parameters, the item parameters, and their difference, are the *logits*.

Rasch model is based on three main assumptions, namely linearity, comparison invariance, and local independence. These assumptions are briefly outlined in the following paragraphs, with a specific focus on conditional independence and on the consequences of its violation.

Linearity of the scores. The linearity of the scores is obtained with the logarithm transformation of the odds. By applying this transformation, person's parameters and item parameters are placed on the same continuous latent latent trait. The measurement units of the latent trait are the *logits*, which define an interval scale for the interpretation of the scores.

This linear transformation allows for setting the lowest parameter observed equal to 0, without losing the original relationship between the estimates. Consequently, comparison invariance (described in the following paragraph) assumption is satisfied.

Comparison invariance. The comparison between any two persons is independent from the set of stimuli on which the comparison is based, as well as independent from the comparison between any other two persons.

The same holds for the stimuli, so that the comparison between any two items is independent from the respondents on which the comparison in made, as well as from the comparison between any other two persons.

The comparisons are invariant in the sense that the comparison between persons only depends on the ability parameters of those two persons, and the comparison between stimuli only depends on the stimuli properties.

Local independence. According to the Rasch model, a person with a high level of ability β will have a greater probability of respond correctly to almost every item. Conversely, a person with a low level of ability β will have a lower probability of responding correctly to almost every item. The variability between the items can hence be explained only in terms of person's level of ability β. As such, ability β can be considered as the source of general dependence between the items, and, once it is accounted for, any relationship between the items should disappear. The capacity of the persons' parameters β to explain all the variability between the responses, and hence the absence of a further relationship between the items, is called local independence (?, ?).

The statistical independence of responses implies that the probability of correctly responding to different items is equal to the product of the probabilities of answering each of them correctly. The local independence of the responses can be formalized as follows:

$$P(\boldsymbol{X}) = \prod_{p} \prod_{s} P(x_{ps}), \tag{1.11}$$

where $\boldsymbol{X}$ is the $p \times s$ matrix of the responses.

Violation of local independence can happen in two main instances, either by involving multidimensionality or response dependence. The consequences of the local independence violation due to either multidimensionality or response dependence move in opposite directions but they both result in less reliable parameters estimates and predictions.

Unidimensionality posits that item responses are explained by only one latent trait dimension, shared by both respondents and stimuli, and it is the basic underlying assumption of the Rasch model. Multidimensionality refers to those cases where there are unexpected person's parameters other than β involved in the responses to the items.

Multidimensionality is indeed a property of many different scales for psychological assessment. For instance, the Big Five Questionnaire (?, ?) is a questionnaire for the assessment of the Big Five personality traits, composed of different subscales. The items in each of the 5 subscales are aimed at assessing one of the 5 personality traits posited by the Big Five theory (i.e., agreeableness, extroversion, openness to experience, neuroticism, conscientiousness), and they can be grouped according to the personality trait they aim for. As such, they show a within–subscale variability which cannot be explained by only the person's parameter β.

Multidimensionality can also raise from stimuli linked by common attributes such as a common item stems, common stimulus materials, or common item structures (?, ?). Consequently, stimuli will display a variability that cannot be understood just in terms of ability parameters beta.

Response dependence (i.e., for a fixed person, hence for a fixed level of ability β, the response to an item might depend to the response to a previous item ?, ?) violates the local independence assumption as formulated in Equation 1.11. The probability associated to the responses to each item are not independent events anymore, hence their probabilities cannot be multiplied. Consequently, it is not possible to state that the probability of correctly responding the entire set of items corresponds to the product of the probabilities of responding correctly to each of them.

Response dependence can happen when the response given to an item is used as a clue for responding to the following item. Response dependence might also arise during the performance at computerized task, when the response to a previous stimulus might leave a carry-over effect on the response to the following stimulus (?, ?). Specifically, in the second case, the variability at the item level is affected by new sources of variability which are mostly composed of error variance.

Violating local independence affects the fit of data to the model (?, ?), and produces unreliable parameter estimates (e.g., ?, ?, ?). When local dependence is due to multidimensionality, extra sources of random noise are added to data. The error variance is hence increased, producing less accurate and reliable predictions. When local dependence is due to response dependence, the similarity of responses of persons across items and responses is higher, leading to a lower error variability in the data.

1.2 Modeling time responses

By modeling response times within an IRT approach, an interaction between the parameters defining persons' accuracy responses and time responses is implicitly assumed. This is nothing else than the speed-accuracy trade-off also reported in analysis of IAT data (e.g., ?, ?).

Traditionally in IRT modeling, the speed-accuracy trade-off has been expressed by adopting a regression parameter for respondents' ability on their response times. Consistently with IRT models in general, also items are fundamental in determining the time responses. It is commonly assumed that more difficult items do need for more time to get a response, so that also an item time parameter is needed (?, ?).

1.2.1 Log-normal model

A log-normal model for the analysis of the response times on a test has been introduced by ? (?). This model is part of a hierarchical model for the modeling of accuracy and time responses given to a test in an IRT framework (?, ?). As also stated by (?, ?, ?), the model used for accuracy responses (that are the IRT model presented in the previous section) and time responses can be used separately for the analysis of the accuracy and time responses to a test.

The advantage of using the hierarchical approach in (?, ?) is that the relationship between IRT and time parameters can be studied and understood at a second level of modeling.

The log-normal model, as its name suggests, assumes a normal density distribution for the logarithm of the time responses. The use of a log-normal family can be traced to its good fit to the observed data already observed in previous work (e.g., ?, ?, ?). More trivially, it comes natural to model with a normal distribution (defined over the entire real continuum) the log transformation of a variable that is a non-negative variable by definition (the response times) (?, ?).

The structure of the original formulation of the log-normal model is analogous to the one of the 2PL IRT model in Equation 1.2 in Section 1.1 for mainly three reasons.

Firstly, both 2PL and log-normal impose the same structure on the mean of the distribution of the binary response variable and on that of the distribution of the continuous variable, respectively. In both cases, the mean is represented by the difference between respondents' and items parameters operating in opposite directions.

Moreover, both models assume a parameter that changes the relationship between the item parameter and the respondent's parameter, namely a discrimination parameter. Further details on the effect and the interpretation of the discrimination parameter are illustrated after the mathematical specification of the log-normal model.

Finally, given the nature of the distribution of response times (i.e., it is bounded at 0), the log-normal model does not need the definition of a lower asymptote (i.e., a guessing parameter like in 3PL model in Equation 1.3).

Response time t (i.e., the realization of a random variable T) of person p on item s can hence be expressed by positing the normal density distribution of the log-response time:

$$f(t_{ps}|\tau_p, \delta_s) = \frac{\alpha_s}{t_{ps}\sqrt{2\pi}} exp\left\{-\frac{1}{2}[\alpha_s(ln\ t_{ps} - (\delta_j - \tau_i))]^2\right\}. \tag{1.12}$$

As in IRT models, both respondents' parameters τ_p and δ_s are allowed to vary between $-\infty$ and $+\infty$. Although the sign of the persons' parameters is reversed, the mean of the distribution of Equation 1.12 resembles the one of the 2PL in Equation 1.2. The change in the sign of respondents' parameters allows for interpreting the parameter as a speed parameter, according to which, the larger the value of τ_p, the faster the responses given across items (i.e., the respondent tend to spend less time on the items).

Parameter δ_s describes the time intensity (or time consumingness) of an item, which is the time the stimulus requires from the respondents. The larger the value of δ_s, the higher the amount of time respondents need to give the response. Parameter α (i.e., the reciprocal of the standard deviation of the standard deviation of the normal distribution) is the discrimination parameter of the model. A larger value of α_s means less dispersion for the log-response time distribution on item s. Consequently, it can be said that the item has a better discrimintating ability between different respondents with different levels of speed. Parameter α affects the relationship between respondents' speed τ_p and item time intensity δ_s, similarly to what happens when parameter a_s changes in the 2PL model. If the value of α_s increases, the distributions of the log-time for any two values for the speed parameter show less overlap.

The 1PL model (Equation 1.1) in Section 1.1 can be considered as a constrained model deriving from the 2PL one (Equation 1.2). The constraint is imposed on the discrimination parameter a, which is forced to be equal across all items (1PL, Rasch model). A similar reasoning can be done for the log-normal model, by either forcing α_s to be equal across all items s ($\alpha_s = \alpha_s$ for all s), or forcing α to be one for all items s ($\alpha_s = 1$ for all s). These constraints bring a parametrization similar to that of the 1PL/Rasch model.

In the empirical application in ? (?), both a non-constrained and a constrained version of the log-normal model were tested in terms of goodness of fit to the data. The normal analogs of the non-constrained and constrained models were tested as well.

The normal models were the ones showing the worst goodness of fit, while constraining α to be equal across all items did not affect much the log-normal models goodness of fit.

1.3 Linear Mixed Effects Models

As illustrated in the previous Sections, an IRT (or Rasch) approach for modeling both accuracy and time responses provide a detailed information on the parameters that determine the observed responses.

The use of a log-normal model can indeed overcome the issue related to the discretization of the response times for the application of the Many Facet Rasch Model in Section **??** of Chapter **??**. The use of a separate model for accuracy responses, always under an IRT or Rasch framework, allows for obtaining useful and detailed information also from accuracy data, with a similar parameterization as that obtained from the log-time responses. Potentially, the estimates obtained from the two models can be combined at a second level of modeling by using hierarchical approach as that illustrated in (?, ?).

Despite this approach sounds promising, it cannot account for the fully-crossed design of IAT data, and the sources of dependency related to it. As thoroughly illustrated in the introduction, the fully-crossed design of the IAT comes with several sources of variability at different levels. These sources of variability generate dependency at the level of the single observations, which violates the assumption of conditional independence. Conditional independence is not only a necessary assumption for the application of the Rasch model, but is a basic assumption needed for obtaining reliable results with any statistical analysis. Violating the assumption of conditional independence brings to biased parameter estimates which can in turn lead to an inflated probability of committing Type I error or in an underestimation of the importance of the experimental condition (?, ?, ?, ?).

Linear Mixed Effects models (LMMs) are the most straightforward way to deal with this data structure. Moreover, both respondents and stimuli can be conceptualized as random factors by specifying the appropriate random structure, hence the issues of the *by-participant* or *by-stimulus* analyses are overcome.

LMMs allow for decomposing the error variance by specifying an appropriate random structure with different levels, that are the levels where uncontrolled random variation can be reasonably found (?, ?).

The error variance can be partitioned into random effects that reflect the assumption on the structure of dependency created by the random variability at the different levels. By doing so, the multilevel structure of the data, which reflects the random variability of the population from which the various levels are drawn, is accounted for (?, ?).

Finally, LMMs can be applied to both continuous data, such as the log-transformation of the response times, and to dichotomous responses. In the latter case, a Generalized Linear Model (GLMMs) is needed, with the appropriate link function expressing the relationship between the linear combination of the predictors (i.e., linear component of the model) and the observed response.

1.3.1 Generalized Linear Mixed Effects Model and Rasch Model

In a Generalized Linear Model (GLM), the linear predictors are not directly related with the observed response. They need to be linked with a specific function, which goes under the name of link function. The type of link function that needs to be used depends on the nature of the observed variables (?, ?).

For the illustration of the structure of the GLM, and of its expansion for including random effects, we focus on the case of binomial responses $x_{ps} \in \{0, 1\}$, describing the accuracy responses at the IAT.

The linear combination of the predictors is defined by the form of the model expressed by the model matrix $\boldsymbol{X}$, and it determines the linear component of the model. The linear component is defined for each cell of the $p \times s$ $\boldsymbol{X}$ matrix, and it is identified with η_{ps}.

The natural link function g that relates observed binomial responses with the linear component of the model η_{ps} is the *logit* (the logarithm of the odds, ?, ?), and it yields a probability value μ_{ps}:

$$\eta_{ps} = logit(\mu_{ps}) = ln\left(\frac{\mu_{ps}}{1-\mu_{ps}}\right), \tag{1.13}$$

where μ_{ps} is the probability associated to each observed response x_{ps}.

Each link function is an invertible function, and the inverse for the *logit* link function is expressed as:

$$\mu_{ps} = logit^{-1}(\eta_{ps}) = \frac{1}{1+exp(-\eta_{ps})}. \tag{1.14}$$

The structure of the inverse *logit* link in Equation 1.14 can be Equated to the Rasch formulation in Equation 1.10, and, consequently, it is possible to obtain a Rasch parametrization of the data by using a GLM on binomial responses with a *logit* link function (?, ?, ?, ?).

From now on, the parameters of the Rasch model will be referred to as θ_p and b_s, referring to respondents' ability and stimuli difficulty, respectively, to distinguish them from the parameters obtained with the LMMs.

When there are reasons to believe that sources of variability can generate dependency between the observations, such as in the IAT case, random effects accounting for the random factors generating the uncontrolled random variability should be included in the model matrix of the linear component. By doing so, the error variance is partitioned in different levels. The levels in which the error variance is partitioned are defined by the factors considered as random. The partition of the error variance into specific factors is what it makes it controllable and accountable for (?, ?).

The $\boldsymbol{X}$ matrix that defines the linear component of the GLM needs to be extended to also include the random factors. The linear component hence takes on the form:

$$\eta = \boldsymbol{X}\beta + \boldsymbol{Z}d, \tag{1.15}$$

where β indicates the coefficients for the fixed effects, $\boldsymbol{X}$ is the model matrix of the fixed effects β, $\boldsymbol{Z}$ is the $p \times q$ matrix of the random effects (i.e., q is the dimension of the random effects vector), and d is the vector of random effects predictors.

The dimension q of d is defined by the number of levels of each random factor, such as “respondents”, “items”, “nations” and so on, making the dimension of d potentially very large (?, ?). The distribution of the random effects is estimated as a multivariate normal distribution (i.e., $\mathcal{MVN}$) with mean 0 and a $q \times q$ variance-covariance matrix $\boldsymbol{\Sigma}$, which is determined by a single vector parameter Γ (?, ?). The dimension of Γ is usually rather small, and its size is determined by the number of random factors considered. For instance, in a model where respondents, items, and the items variability in three different conditions are specified as random factors, the dimension of the vector parameter Γ is 5. The dimension of the vector parameter Γ remains 5, regardless of the number of respondents or stimuli used.

The objective of LMMS is then to estimate the parameters of the fixed effects as defined by vector β and the parameters of the random effects, defined by vector Γ. Consequently, the parameters estimated for the random factors are not the parameters associated to each level of each factor, but the variance of the populations from which the random factors are drawn. This is the reason why d is indicated with a Latin letter, because it does not indicate population parameters.

Nonetheless, a measure for each level of each random factor is obtained in the form of *conditional modes*, which are the values that maximize the conditional density of the random effects given the vector of parameters (fixed and random) and the observed data (?, ?). The conditional modes that describes the deviation from the fixed factors of each level of each random factors are meta-parameters (?, ?), and are usually referred to as *Best Lineaer Unbiased Predictors* (BLUP, ?, ?).

BLUP are used for the estimation of the Rasch model parameters.

Concerning the stimuli, the easiness estimates b_s are obtained by adding the conditional mode of each stimulus, considered as a random factor, to the estimates of the fixed effects. In the IAT case, the higher the value of stimuli easiness b_s, the easier the stimulus, meaning that it is easily recognized and sorted to the category to which it belongs.

Similarly, adding the conditional mode of each respondent to the estimates of the fixed effects results in the respondents' ability estimates θ_p. In the IAT case, the higher the value of θ_p, the higher the ability of the respondent, meaning that he/she correctly categorizes the majority of the stimuli.

In the Rasch model, respondents' ability and stimuli easiness move in opposite directions. When respondents' ability and stimuli easiness are obtained by using GLMMS, their estimates move in the same direction, hence resulting in an additive effect. The item parameter b_s can no longer be interpreted as an impediment property (difficulty) of the item but it should be interpreted as a facilitation property of the stimulus (easiness) (?, ?, ?). When both θ_p and b_s are high, then the probability of a correct response is higher. When high values of θ_p are combined with low values of b_s, the probability of a correct response for each respondent is as much penalized as their ability cannot balance out the stimulus easiness.

Log-normal model estimates. Rasch model estimates can be obtained by combining together the fixed and random components of GLMMs applied on accuracy responses. Instead of being governed by the difference between respondents' ability and stimuli easiness, the probability of a correct response is governed by the additive effect of respondents' ability and item easiness. Consequently, the interpretation of the stimuli parameters δ_s, changes.

In a similar vein, log-normal model estimates can be obtained by combining the fixed factors to the random factors of LMMs applied to log-time responses. In the typical formulation of the log-normal model (Equation 1.12), the mean of the distribution of the expected log-time responses is expressed by the difference between stimuli time intensity parameters δ_s and respondents' speed parameter τ_p (i.e., $\delta_s - \tau_p$). In the LMMs, the mean of the distribution is defined by the additive effect between respondents' and stimuli characteristics, which move in the same direction. Consistently, the lower the value of speed parameter τ_p, the higher the speed, and the lower the value of δ_p, the lower the time each stimulus require for getting a response.

When respondents with a low value of τ_p (i.e., high speed) respond to items with a low δ_s (i.e., low time intensity), the response times will be faster.

1.4 Random structures

The random structures of the GLMMs and that of the LMMs are the same. The features differentiating the models are the assumptions on the error term ε and the dependent variable. In the GLMMs, the error term is supposed to follow a logistic distribution (i.e., $\varepsilon \sim \mathcal{L}(0, \sigma^2)$, where $\mathcal{L}$ is used to denote the logistic distribution of the disturbance an in ? (?)) and the dependent variable is the accuracy response to each trial of the IAT. In the LMMs, the error term is supposed to follow a normal distribution (i.e., $\varepsilon \sim \mathcal{N}(0, \sigma^2)$), and the dependent variable y is the log transformation of the time response to each trial of the IAT, regardless of whether the answer is correct or not.

The expected response y for each observation $i = \{1, \dots, n\}$ for participant $p = \{1, \dots, P\}$ on stimulus $s = \{1, \dots, S\}$ in condition $c = \{1, \dots, C\}$ can hence be either the *log-odds* of the probability of a correct response (GLMMs) or the log-time of the response (LMMs).

In both GLMMs and LMMs, the fixed intercept α is set at 0. IAT associative conditions c are specified as fixed effect $\beta_c X_c$. Since the intercept is set at 0, none of the level of the fixed effect is taken as reference value. Consequently, the marginal *log-odds* of a correct response for each condition (GLMMs) and the marginal average log-time for each condition (LMMs) are estimated. The fixed part of the models is kept constant, only the random structures change across models.

GLMMs applied on accuracy responses are identified by a capital “A”. LMMs applied on log-time responses are identified by a capital “T”.

1.4.1 Generalized Linear Mixed Effect Models

Model A1 presents the simplest random structure, where only the between–respondents across–conditions variability and the between–stimuli across–conditions variability are considered by specifying both respondents and stimuli as random intercepts across associative conditions:

$$y_i = logit^{-1}(\alpha + \beta_c X_c + \alpha_{p[i]} + \alpha_{s[i]} + \varepsilon_i), \tag{1.16}$$

with

$$\alpha_p \sim \mathcal{N}(0, \sigma^2_{\alpha_p}) \text{ and } \alpha_s \sim \mathcal{N}(0, \sigma^2_{\alpha_s}) \tag{1.17}$$

Model A1 should be preferred when a low within–respondents between–conditions variability, as well as a low within–stimuli between–conditions variability, are observed. This lack of variability at both respondents and stimuli levels might indicate a lack of the IAT effect at both levels. Nonetheless, the random structure of Model A1 results in the estimation of overall respondents' ability estimates θ_p and overall stimuli easiness estimates b_s.

Respondents' ability estimates inform about the overall ability they showed in performing the categorization task, and they can be used as a measure of individual differences for further analysis.

Stimuli overall easiness estimates provide information on the stimuli functioning in respect to their own category. Stimuli belonging to the same category are supposed to be prototypical exemplars of that specific category, and, as such, to be easily recognized and correctly assigned to their category. Consequently, they should have similar easiness estimates. If a stimulus is not recognized as a prototypical exemplar of its alleged category, it will have a higher chance of getting incorrect responses (i.e., being assigned to the incorrect category), from which a lower easiness estimate follows. By comparing the easiness estimates of the stimuli belonging to the same category between each other, it is possible to investigate whether the stimuli belonging to the same category are all easily recognizable as prototypical exemplars or not.

Model A2 accounts for the within–stimuli between–conditions variability and the between–respondents across–conditions variability. Stimuli are specified as random slopes in the associative conditions, respondents are specified as random intercepts across associative conditions, as follows:

$$y_i = logit^{-1}(\alpha + \beta_c X_c + \alpha_{p[i]} + \beta_{s[i]} c_i + \varepsilon_i), \tag{1.18}$$

with:

$$\beta_{sc} \sim \mathcal{MVN}(\mathbf{0}, \mathbf{\Sigma}_{sc}) \tag{1.19}$$

$$\alpha_p \sim \mathcal{N}(0, \sigma^2_{\alpha_p}), \tag{1.20}$$

where $\mathbf{\Sigma}_{sc}$ represents the variance-covariance matrix of the population of the stimuli. It expresses the by-stimulus variability in the associative conditions. The higher the covariance of the stimuli in the two conditions, the more similar is their functioning in the two conditions. Model A2 should result as the best fitting model when a high within–stimuli between–conditions variability is observed and respondents have a low between–conditions variability. This model results in condition–specific stimuli easiness estimates b_{sc} and overall ability estimates θ_p.

The low variability at the respondents' level might already indicate a lack of the IAT effect on their accuracy performance (i.e., ability remains constant across conditions). In other words, the IAT associative condition do not have an effect on the respondents' ability to sort stimuli. As for the overall ability estimates obtained with Model A1, these estimates can be used as a measure of individual differences in performing the categorization task for further analysis.

Conversely, the high within–stimuli between–conditions variability indicate that the stimuli functioning is in some way affected by the specific associative condition and that stimuli characteristics (i.e., the category to which they belong) make them more easily categorizable in one condition than in the opposite one. Thus, condition–specific stimuli easiness estimates allow for investigating whether stimuli functioning differs between conditions.

Consider a stimulus representing a can of coke in the Coke IAT example. If the stimulus presents a higher easiness estimate in the Coke-Good/Pepsi-Bad condition than in the opposite one, it implies that it was more easily sorted when it shared the response key with *Good* rather than *Bad* attributes. Consequently, the differential measures computed on the condition–specific stimuli estimates inform about the contribution of each stimulus to the IAT effect, which in turn lead to a better understanding of the automatic associations driving the effect.

The random structure of Model A3 has the same level of complexity as that of Model A2. However, the multidimensionality of the error term is specified for the respondents and not for the stimuli. Model A3 accounts for the within–respondents between–conditions variability and between–stimuli across–conditions variability by specifying respondents as random slopes in the associative conditions and stimuli as random intercepts across associative conditions:

$$y_i = logit^{-1}(\alpha + \beta_c X_c + \alpha_{s[i]} + \beta_{p[i]} c_i + \epsilon_i), \tag{1.21}$$

with:

$$\beta_{pc} \sim \mathcal{MVN}(\mathbf{0}, \mathbf{\Sigma}_{pc}) \tag{1.22}$$

$$\alpha_s \sim \mathcal{N}(0, \alpha_s^2), \tag{1.23}$$

where $\mathbf{\Sigma}_{pc}$ represents the variance-covariance matrix of the population of the respondents. It expresses the by-respondent variability according to the associative conditions. The high covariance does not necessarily implies that the performance is not affected by the associative condition. For instance, an high ability respondent might be an high ability respondent in both conditions, although his performance might be affected by the associative conditions. Model A3 should result as the best fitting model when a low within–stimuli between–conditions variability and a high within–respondents between–conditions variability are observed. This model results in condition–specific respondents' ability estimates θ_{pc} and overall easiness estimates b_s.

As in Model A1, the lack of within–stimuli between–conditions variability might indicate that the stimuli functioning is not affected by the associative condition in which they are presented. The overall easiness estimates can still inform about the stimuli functioning in respect to their own category.

The high within–respondents between–conditions variability at the respondents level indicate that the IAT associative conditions affect the accuracy performance of the respondents, or, in other words, that their ability level is in some way hindered by one of the associative conditions. A measure of the bias due to the associative conditions can be obtained by computing the difference between each respondent condition–specific ability estimate.

A summary of the Rasch model estimates that can be obtained from the three random structures is reported in Table 1.2

Table 1.2: Rasch model and log-normal model estimates.

	Rasch model		Log-normal model	
Model	Respondents	Stimuli	Respondents	Stimuli
1	Overall (θ_p)	Overall (b_s)	Overall (τ_p)	Overall (δ_s)
2	Overall (θ_p)	Condition–specific (b_{sc})	Overall (τ_p)	Condition–specific (δ_{sc})
3	Condition–specific (θ_{pc})	Overall (b_s)	Condition–specific (τ_{pc})	Overall (δ_s)

Note: Respondent $p = 1, \ldots, p$, Stimulus $s = 1, \ldots, S$, Condition $c = 1, \ldots, C$, where P, S, and C, are the number of respondents, stimuli, and conditions, respectively.

1.4.2 Linear Mixed Effect Models

Model T1 presents the simplest random structure. Only the between–respondents across-conditions variability and the between–stimuli across–conditions variability are considered by specifying both respondents and stimuli as random intercepts across associative conditions:

$$y_i = \alpha + \beta_c X_c + \alpha_{p[i]} + \alpha_{s[i]} + \varepsilon_i, \tag{1.24}$$

with

$$\alpha_p \sim \mathcal{N}(0, \sigma^2_{\alpha_p}) \text{ and } \alpha_s \sim \mathcal{N}(0, \sigma^2_{\alpha_s}) \tag{1.25}$$

Model T1 should be preferred when a low within–respondents between–conditions variability and a low within–stimuli between–conditions variability are observed. The lack of variability at both respondents and stimuli levels might indicate that there is no IAT effect at both levels. Model T1 allows for estimating overall respondents' speed estimates τ_p and overall stimuli time intensity estimates δ_s.

Respondents' speed estimates inform about the overall speed with which they have been performing the categorization task. As the ability estimates obtained from Model A1, the overall speed estimates can be used as a measure of individual differences in further analysis.

As overall easiness estimates, stimuli overall time intensity estimates inform about the stimuli functioning in respect to their own category. If the stimuli belonging to the same category are equally recognized as prototypical exemplars of their category, they should require a similar amount of time for getting a response, and hence they should have a similar time intensity. If a stimulus presents characteristics that make it less recognizable as prototypical of a specific category (e.g., a picture of a can of soda that is not immediately recognizable as either Coke or Pepsi), it might require more time for being identified and sorted. Consequently, it should have a higher time intensity estimate. By comparing the easiness estimates of the stimuli belonging to the same category between each other, it is possible to investigate whether the stimuli belonging to the same category require a similar time for getting a response. In doing so, other stimuli characteristics should be taken into account, especially when attributes stimuli are considered. While mages stimuli are almost immediately processed and sorted, attribute stimuli need to be read and understood before they can be assigned to a category. Consequently, the familiarity with a specific term might play an import role in its recognition and sorting, hence positively (if it is a familiar term) or negatively (if it is an unfamiliar term) affecting its time intensity. Also the length of the word itself might influence stimuli time intensity.

Model T2 accounts for the within–stimuli between–conditions variability and the between–respondents across–conditions variability. The random slopes of the stimuli in the associative conditions and the random intercepts of the respondents across associative conditions are specified:

$$y_i = \alpha + \beta_c X_c + \alpha_{p[i]} + \beta_{s[i]} c_i + \varepsilon_i, \tag{1.26}$$

with:

$$\beta_{sc} \sim \mathcal{MVN}(\mathbf{0}, \mathbf{\Sigma}_{sc}) \tag{1.27}$$

$$\alpha_p \sim \mathcal{N}(0, \sigma^2_{\alpha_p}), \tag{1.28}$$

where $\mathbf{\Sigma}_{sc}$ represents the variance-covariance matrix of the population of the stimuli, and it expresses the by-stimuli variation according to the associative condition. As for accuracy models, the higher the covariance, the more similar the stimuli functioning in the two conditions.

Model T2 should result as the best fitting model when a high within–stimuli between–conditions variability is observed and respondents have a low between–conditions variability. This model results in condition–specific stimuli time intensity estimates δ_{sc} and overall speed estimates τ_p.

The low variability at the respondents' level might already indicate a lack of the IAT effect on their speed performance (i.e., speed remains constant across conditions). In other words, respondents are not adjusting their speed to the specific associative condition. As for the overall speed estimates obtained with Model T1, these estimates can be used as a measure of individual differences in performing the categorization task for further analysis.

Conversely, the high within–stimuli between–conditions variability indicate that the stimuli do require a different amount of time to be sorted according to the associative condition in which they are presented. Their functioning is hence affected by the associative conditions, and the condition–specific time intensity allow for investigating how and how much.

The differential measure computed between the condition–specific time intensity estimates provide a measure of the bias on the time each stimulus require for getting a response due to the associative conditions. Consequently, the contribution of each stimulus to the IAT effect can be investigated.

The random structure of Model T3 has the same level of complexity as that of Model T2. However, the multidimensionality of the error term is specified for the respondents and not for the stimuli. Model T3 accounts for the within–respondents between–conditions variability and between–stimuli across–conditions variability by specifying the respondents as random slopes in the associative conditions and the stimuli as random intercepts across associative conditions:

$$y_i = \alpha + \beta_c X_c + \alpha_{k[i]} + \beta_{j[i]} l_i + \varepsilon_i, \tag{1.29}$$

with:

$$\beta_{pc} \sim \mathcal{MVN}(\mathbf{0}, \mathbf{\Sigma}_{pc}), \tag{1.30}$$

$$\alpha_s \sim \mathcal{N}(0, \alpha_s^2), \tag{1.31}$$

where $\mathbf{\Sigma}_{pc}$ is the variance-covariance matrix of the population of the respondents, and it expresses the by-respondents variability according to the associative condition. Similarly to accuracy models, a high covariance does not imply that respondents' performance is not affected by the associative conditions but that their baseline speed is making them having a similar performance in both conditions. Model T3 should result as the best fitting model when a low within–stimuli between–conditions variability and a high within–respondents between–conditions variability are observed. This model results in condition–specific respondents' speed estimates τ_{pc} and overall time intensity estimates δ_s.

As in Model T1, the lack of within–stimuli between–conditions variability might indicate that the stimuli functioning is not affected by the associative condition in which they are presented. The overall time intensity estimates can still inform about the stimuli functioning in respect to their own category.

The high within–respondents between–conditions variability at the respondents level indicate that the IAT associative conditions affect the speed performance of the respondents, or, in other words, that their speed is lower in one of the associative conditions. A measure of the bias due to the associative conditions can be obtained by computing the difference between each respondent condition–specific speed estimate.

1.5 Other random structures

The random structures presented in the previous sections are just some of the possible random structures that can be specified for analyzing IAT data. Indeed, since IAT data has a specific structure (illustrated in Section **??** of Chapter **??**), a model with a random structure that decomposes error variance into each of the sources of variation can be specified (Maximal Model, MM; ?, ?).

In the MM, both between–respondents across–conditions variability and within–respondents between–conditions variability is accounted for by specifying respondents random intercept across conditions and their random slopes in the associative conditions. The same can be done for the stimuli, so that they are specified as random intercepts across conditions and as random slopes in the associative conditions. Moreover, the variability due to the interaction between the stimuli and the respondents variability (i.e., respondents' individual reactions to each stimulus) can be accounted for by specifying the interaction effect between respondents and stimuli random intercepts.

MM results in the estimation of the weights associated with each level of the fixed effect, as well as in the estimation of the variance of the population to which each factor considered as random belongs. In this case, the stimuli, the respondents, and their interaction. This interaction can be considered as the variability due to the idiosyncratic reactions of each respondent to each stimulus. Also the variance-covariance matrix for each level on which the multidimensionality of the error variance was allowed are estimated. Therefore, the variance of the respondents in each level of the associative condition variable, as well as their covariance, are estimated. The same is done for the stimuli.

By considering two levels of the fixed effect of the associative conditions and removing the intercept by setting it at 0, this model results in the estimation of 18 parameters, two of which are the weights of the fixed effects. Three parameters are referring to the estimated variances of the respondents and stimuli population, and the interaction between them. Three parameters are estimated for the multidimensionality of the associative conditions on the respondents (the variance in the two conditions and their covariance), as well as three parameters for the multidimensionality of the associative conditions on the stimuli (the variance in the two conditions and their covariance). Finally, one parameter refers to the estimated residual variance.

A model of such a complexity needs an extremely high variability at each level of the random structure to converge. Beyond being at risk of convergence failure, it is also at risk of over-fitting the data (?, ?), hence resulting in biased and not-interpretable estimates.

A model with the random structure of the MM is neither needed nor appropriate for the estimation of the Rasch model and log-normal model estimates from IAT data. By specifying both respondents and stimuli as random intercepts and random slopes in the associative conditions, overall and condition–specific estimates can be obtained for each factor. The difference between each of the condition specific estimates and the overall estimates provides information about the bias due to each condition on either the respondents or the stimuli. The difference between condition–estimates results in a measure of the bias due to the IAT associative conditions. Consequently, it allows for investigating the impact of the IAT associative conditions on either respondents' performance or stimuli functioning. When the IAT is used, the focus is usually on this difference, expressing the IAT effect. Therefore, the estimation of the overall estimates for both respondents and stimuli can be dropped, without losing important information.

For the Rasch model or the log-normal model to be identified, either respondents or stimuli have to be centered around 0 (e.g., ?, ?). This can be done by setting the fixed intercept at 0 and by specifying either respondents or stimuli as random variation (i.e., random intercepts) around it. As such, each respondents or stimuli BLUP defines the deviation of each level of the considered factor from 0, that is, the average of the respondents or stimuli estimates. Consequently, only either respondents or stimuli can be specified as random slopes in the associative conditions, while the other must be specified as random intercepts. The decision on where to allow for the multidimensionality of the associative condition, whether on the respondents or the stimuli, should be driven by the observed variability in the data.

Finally, the estimation of the interaction effect between stimuli and respondents random intercepts do require an high respondents $\times$ stimuli variability to avoid convergence failure. Consequently, it can be dropped and added to the model only in those cases in which the error variance is still high after the estimation of all the other parameters (?, ?, ?).

Clearly, also other fixed effects could have been included in the model. For instance, also the belonging category of the stimuli, which is indeed an indepedent variable as illustrated in Section **??** of Chapter **??**, could have been included as a fixed effect. However, we decided to focus the attention on the effect of the IAT associative condition, and on the deviation from it of each of the level of either the stimuli or the respondents. In our opinion, the information yielded from a model with this structure is more useful in gaining insights on the IAT functioning, for example by highlighting the stimuli giving the highest contribution to the IAT effect. Indeed, by specifying the fixed effect of the stimuli categories, an information of the respondents' or stimuli deviations from their mean could have been obtained. However, while this information is useful and meaningful for the stimuli functioning, it is not that meaningful and useful for the respondents. What does it mean that respondent p has an impairment of 1.06 on stimulus *pain* of *Bad* category? This information might be more useful if also the interaction between the stimuli categories and the associative conditions is specified. However, this interaction would need an extremely high variability for the model to converge and to provide meaningful estimates.

We decided to keep a more parsimonious model by including as a fixed effect a factor that would have provided useful information regarding both respondents and stimuli, namely, the associative condition. Nothing is preventing anyone for including other fixed effects, and to investigate whether the model does converge or not. Since the aim of the thesis was to provide a general modeling framework for implicit measures data, we decided to go for a more parsimonious but generalizable model.

Finally, considering only the fixed effect of the condition, hence allowing for the multidimensionality only according to this effect, is in line with previous applications of the Rasch model to IAT data (e.g., ?, ?).