A guided tutorial on linear mixed-effects models for the analysis of accuracies and response times in experiments with fully-crossed design

Ottavia M. Epifania[1], Pasquale Anselmi[1], and Egidio Robusto[1]

[1] Department of Philosophy, Sociology, Education, and Applied Psychology, University of Padova

Ottavia M. Epifania https://orcid.org/0000-0001-8552-568X

Pasquale Anselmi https://orcid.org/0000-0003-2982-7178

Egidio Robusto https://orcid.org/0000-0002-7583-2587

The authors have no conflict of interest to disclose.

Author Notes and Acknowledgments:

The authors would like to thank Professor Michele Vicovaro and Professor Mario Dalmaso for sharing the data, and Professor Livio Finos for his insightful suggestions during the revision of the manuscript.

Correspondence concerning this article should be addressed to: Pasquale Anselmi, Department of Philosophy, Sociology, Education, and Applied Psychology, Via Venezia 14, Padova, Italy.

E-Mail: pasquale.anselmi@unipd.it

Abstract

Experiments with fully-crossed designs are often used in experimental psychology spanning several fields, from cognitive psychology to social cognition. These experiments consist in the presentation of stimuli representing super-ordinate categories, which have to be sorted into the correct category in two contrasting conditions. This tutorial presents a linear mixed-effects model approach for obtaining Rasch-like parametrizations of response times and accuracies of fully-crossed design data. The modeling framework for the analysis of fully-crossed design data is outlined along with a step-by-step guide of its application, which is further illustrated with two practical examples based on empirical data. The first example regards a cognitive psychology experiment and pertains to the evaluation of a spatial-numerical association of response codes effect. The second one is based on a social cognition experiment for the implicit evaluation of racial attitudes. A fully-commented R script for reproducing the analyses illustrated in the examples is made available as supplementary material.

Keywords: Mixed-effects models; Fully-crossed data design; Rasch model; Log-normal model; R software

Introduction

In experiments with fully-crossed design, each respondent is presented with the same set of stimuli in two contrasting conditions, and their task consists in sorting those stimuli into the correct super-ordinate category that they represent (see, e.g., Westfall et al., 2014). These experiments are widely used in different fields, such as cognitive psychology and social cognition. For instance, in cognitive psychology, they can be used for the evaluation of the Stroop effect or the spatial-numerical association of response codes effect (i.e., SNARC effect, see, e.g., Dehaene et al., 1993; Zhang et al., 2022), while in social cognition they can be used for the implicit investigation of attitudes and preferences (see, e.g., Epifania et al., 2022b). Examples of the latter can be the implicit association test (IAT; Greenwald et al., 1998) or the Sorting Paired Features task (SPF; Bar-Anan et al., 2009). This tutorial presents a linear mixed-effects models (LMMs) approach for the analysis of fully-crossed design data, which is appropriate for obtaining a Rasch-like parametrization of response times and accuracies. The Rasch-like parametrization results in fine-grained information at respondent and stimulus levels, which is needed for a deeper understanding of the functioning of these experiments. For instance, it allows for the investigation of the contribution of each stimulus to the effects evaluated in these experiments (see, e.g., Epifania et al., 2022a). Two applications on real data from cognitive psychology (i.e., an experiment entailing a SNARC-like effect) and social cognition (i.e., data from an IAT) are illustrated.

Both De Boeck et al. (2011) and Doran, Bates, Bliese, and Dowling (2007) described the application of Generalized LMMs (GLMMs) for obtaining a Rasch-like parametrization of response accuracies. The approach presented in this tutorial extends that by the authors to experiments with fully-crossed designs, as well as to the analysis of the log-time responses. Examples of applications of LMMs for the analysis of fully-crossed design data can be found in Wolsiefer, Westfall, and Judd (2017). While Wolsiefer et al. (2017) model the contrasts between the conditions and the stimulus type as well as their interaction, this tutorial focuses on the effect of the conditions on the variability at either the stimulus level or the respondent level, without considering the contrasts between conditions or stimulus type. Consistent with

the definition in Westfall et al. (2014) and Wolsiefer et al. (2017), the expression "fully-crossed design" is here used to denote data structures obtained from experiments where the same set of stimuli is presented multiple times to all the respondents in two different conditions.

In the following section, fully-crossed design experiments are introduced along with the issues related to the methods that are usually employed for scoring the data. Then, the relationship between generalized linear models and the Rasch model is illustrated, as well as the relationship between the log-normal model (van der Linden, 2006) and the linear model. The codes for estimating the models in R with the `lme4` package (Bates, Machler, et al., 2015) follow, along with two practical applications of the modeling approach on real data. A brief discussion concludes the argumentation.

Fully-Crossed Data Design

In experiments with fully-crossed design, such as those carried out with the IAT or those for the investigation of the SNARC effect, the respondents have to sort a set of stimuli according to a rule. The stimuli represent the super-ordinate categories to which they belong, while the sorting task is performed in two experimental conditions. For instance, consider an experiment for the evaluation of the SNARC effect (i.e., the automatic association between the location of the response button on the keyboard and the magnitude of the stimuli, generally numbers, Dehaene et al., 1993; Zhang et al., 2022), and an IAT for the indirect evaluation of the attitudes towards Black and White people. In the former one, the super-ordinate categories are the two categories that define the magnitude of the numbers, namely "high numbers" and "low numbers". In the latter one, the super-ordinate categories are the two attribute categories, Good and Bad, and the two object categories, White people and Black people. One of the conditions of the experiment is supposed to be "easier" than the other in terms of faster and more accurate responses. This condition is said to be the condition compatible with the automatic association of the respondent. The contrasting condition is supposed to be against the automatic association of the respondents, such that they might struggle in assigning the stimuli to the correct category. The condition might be perceived as more difficult, thus resulting in slower and less accurate responses. Usually, the effect of the contrasting conditions on the performance of the

respondents (i.e., the within–respondent between–condition variability) is the main focus of the investigation. For instance, when scoring the IAT, the difference in the performance of the respondents between the two conditions is know as *IAT effect* (Greenwald et al., 1998, 2003) and an ad hoc effect size measure is employed for expressing its strength and direction (see, e.g., Epifania et al., 2020a; Greenwald et al., 2003).

Figure 1 provides an example of a fully-crossed structure where three respondents (p_1, p_2, p_3) respond to two stimuli (☺, 😷), each presented twice in two contrasting conditions (dashed and solid rectangles). In a multilevel modeling perspective, the respondents (on the left side of the figure) and the stimuli (on the right side of the figure) can be considered at the same level (Judd et al., 2012). Moreover, since each stimulus is presented to each respondent (in this case, each stimulus is presented four times to each respondent across the two conditions), it can be said that the respondents and the stimuli are crossed between each other. The contrasting conditions (i.e., the conditions within which the stimuli are presented to the respondents) are at the highest level, which includes both the respondents and the stimuli. Since the same stimuli are presented to the same respondents in the two contrasting conditions, the respondents and the stimuli are also crossed with the contrasting conditions at the highest level. The trials (represented by the squares between the respondents and the stimuli in the figure) are at the lowest level of observation resulting from the crossing between the respondents, the stimuli, and the conditions. In other words, each trial is a unique combination respondent $\times$ stimulus $\times$ condition.

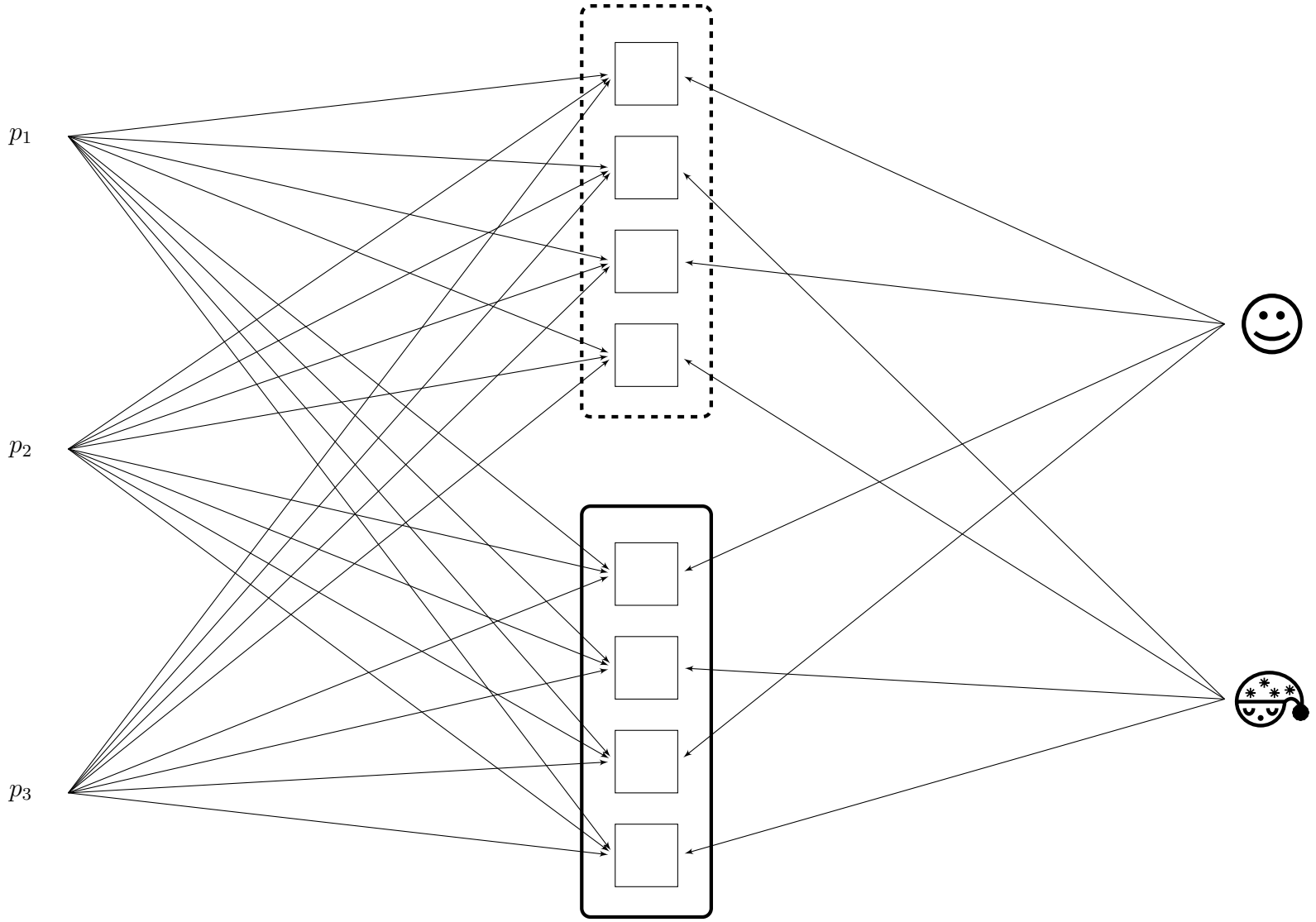


Figure 1: Fully-crossed structure of an experiment where three respondents (p_1, p_2, p_3) respond to two stimuli (☺, 🤒) each presented twice in two conditions (dashed and solid rectangles). Each condition is composed of four trials (the squares).

Data with fully-crossed structures present different sources of variability that should be carefully considered to obtain reliable results (see, e.g., Barr et al., 2013). To better exemplify how the variability arises from fully-crossed structures, consider two respondents, Jane Doe and John Doe, who are presented with a task for the investigation of the SNARC effect. Jane Doe works as an accountant while John Doe is a writer. At the very least, differences are expected in their overall performance (e.g., average response times and number of correct responses across conditions), which can be considered as the expression of the individual differences of the respondents (i.e., between–respondent variability). Due to the nature of her job, Jane is more accustomed to handling numbers than John. As a result, she outperforms John across conditions (i.e., she shows an overall better performance). The sets of exemplars chosen to represent the categories of small numbers (e.g., 0 to 4) and large numbers (e.g., 6 to 9) might present their own variability as well. For instance, the number "0" might be immediately recognized as representing small numbers, as well as the number "9" might be immediately recognized as representing large numbers, while number "4" or number "6" might not be immediately recognized as belonging to the first and second categories, respectively. The between–stimulus variability represents the sampling variability of the stimuli chosen for representing the super-ordinate categories "small numbers" and "large numbers". As the variability observed among respondents primarily stems from their individual differences, the variability between stimuli is largely attributed to the characteristics of the stimuli.

So far, the effect of the contrasting conditions has been considered neither at the respondent level nor at the stimulus level, although their effect on the performance of the respondents is usually the main interest when experiments like these are carried out. The individual differences of the respondents might be exacerbated or diminished by the effect of the contrasting conditions. Despite Jane generally performed better than John, the differences in their performance might be attenuated in the compatible condition (i.e., the condition where small numbers are sorted with the left key and high numbers are sorted with the right key). Being familiar with numbers, Jane might have no issues in performing at the task, while John might actually be favored by the condition that is most "natural" given the SNARC effect (i.e., the compatible condition). However, in the incompatible condition (i.e., the condition where small numbers

are sorted with the right key and high numbers are sorted with the left key), John might struggle more than Jane in sorting large numbers on the left side and small number on the right side. While in the compatible condition the differences between their performance are minimized, they might be exacerbated in the incompatible condition. The effect of the experimental conditions can be investigated by considering the within–respondent between–condition variability, which is the expression of the variability in the performance of each respondent due to the effect of the experimental manipulation.

Also the functioning of the stimuli might be affected by the effect of the experimental conditions, such that some of the stimuli can be more or less easily sorted in one condition than in the opposite one. The within–stimulus between–condition variability expresses the variability in the functioning of the stimuli that is ascribable to the experimental conditions (Epifania et al., 2022a, 2021). For instance, respondents may have specific reactions to the stimuli in the contrasting conditions, such that a specific variability due to the interaction of the performance of each respondent with the functioning of each stimulus should be expected. To best exemplify this situation, consider again the IAT for the investigation of attitudes towards White and Black people. Some of the stimuli used for representing the super-ordinate categories might have peculiarities that could elicit specific reactions in some of the respondents but not in others. For instance, the word "grace" is often used to represent the Good category. However, this word may also represent the proper name "Grace", such that if a respondent has a personal relationship with a person named "Grace", they might have peculiar reactions to this stimulus.

Usual Scoring of Fully-Crossed Data

Usually, to investigate the effect of the contrasting conditions on the performance of the respondents, a standardized measure of the difference in their responses between conditions is computed, which can be considered as an effect size measure. A single score for each respondent is obtained, which expresses the strength and direction of the bias towards one of the two conditions. This approach is usually referred to as by-participant approach, given that the focus is on the performance of the respondents across the stimuli and between the conditions (Judd et al., 2012). Easy-to-compute and easy-to-interpret, effect size measures are quite spread to in-

terpret the results of fully-crossed design data (e.g., the *D* score for scoring the IAT; Epifania et al., 2020a, 2020b; Greenwald et al., 2003). However, two potential pitfalls can be encountered by following this approach, both of which are ascribable to the random variability of the stimuli that is not properly accounted for. Firstly, the information that can be gathered at the stimulus level by acknowledging their variability (i.e., the variability of the stimuli due their multiple presentations within and between conditions) is often overlooked, although a fine-grained analysis at this level might further inform on the psychological processes involved in the task (e.g., it might allow for identifying potentially malfunctioning stimuli or highlighting the stimuli that mostly contribute to the effect of interest; Epifania et al., 2021, 2022a; Wolsiefer et al., 2017). Secondly, if the variability of the stimuli is not accounted for, it can be potentially confounded with the variation due to the contrasting conditions (Barr et al., 2013), such that the resulting scores might include sources of random variability (i.e., error variance) that are not related to the construct of interest. Consequently, any significant difference between the contrasting conditions might be due to the variability at the stimulus level, this inflating the probability of committing Type I error (Barr et al., 2013; Judd et al., 2012; McCullagh & Nelder, 1989).

While the first pitfall can be dealt with an approach focused on item information, such as the Rasch model (Rasch, 1960), the second one can be easily dealt with by using linear mixed-effects models (LMMs). Taking into account the variability related to the fully-crossed structure of the data, LMMs allow for investigating the effect of the experimental manipulation (i.e., the contrasting conditions, considered as fixed factors) while accounting for the random variability of the respondents and the stimuli (considered as random factors). If generalized LMMs (GLMMs) are used for modeling response accuracies, a Rasch-like parametrization of the data can be obtained. Conversely, the application of LMMs to the log-transformed response times results in a parametrization similar to that of the log-normal model (van der Linden, 2006). Finally, LMMs allow for modeling separately accuracies and response times by applying GLMMs and LMMs independently from one another.

Rasch-Like and Log-Normal Parametrizations from Linear Mixed-Effects Models

This section first presents the Rasch model and its similarities with a generalized linear model for binomially distributed responses (GLM), along with the log-normal model and its similarities with the linear model (LM). A section on the random structures to include in the (G)LM to account for the random variability due to different factors (considered as random effects) follows. Finally, the random structures that can be used to obtain Rasch-like parametrization from response times and accuracies of fully-crossed data are illustrated.

Rasch Model, Log-Normal Model, and (Generalized) Linear Models

According to the Rasch model (Rasch, 1960, Equation 1 in Table 1), the probability of respondent p to give a correct response to stimulus s depends on both a respondent's characteristic (i.e., the amount of latent trait of respondent p as described by the ability parameter θ_p) and a stimulus characteristic (i.e., the amount of latent trait needed for giving the correct response to s, as described by the difficulty parameter b_s). In this application, the notation typical of Item Response Theory is used to represent the respondent and stimulus characteristics instead of the typical Rasch notation to avoid a clash of notation with the parameters of LMMs. The higher the value of θ_p, the higher the ability of respondent p (i.e., the higher the number of stimuli correctly endorsed by respondent p). The higher the value of b_s, the higher the difficulty of stimulus s (i.e., the higher the number of incorrect responses observed on stimulus s). In a GLM for binomially distributed responses, the probability μ_{ps} of respondent p to give a correct response to stimulus s is yielded by the inverse of the *logit* link function (i.e., $logit^{-1}$, Equation 2 in Table 1). The *logit* link function connects the observed response accuracies with the linear combination of predictors η_{ps}. The Rasch model (Equation 1) can be equated to the inverse of the *logit* link function $logit^{-1}$ of the GLM.

The log-normal model (van der Linden, 2006, Equation 3 in Table 1) allows for obtaining a Rasch parametrization of log-time responses by considering the observed log-time response as a function of respondent speed (as described by the parameter τ_p) and stimulus time intensity (as described by the parameter δ_s). The larger the value of speed τ_p, the smaller the amount of time respondent p needs for performing the task. The larger the value of time intensity δ_s,

Table 1
Rasch, Log-normal, and Linear models.

Typical	(Generalized) Linear models
Rasch model	
$P(x_{ps} = 1\|\theta_p, \delta_s) = \frac{\exp(\theta_p - b_s)}{1 + \exp(\theta_p - b_s)}$ (1)	$\mu_{ps} = logit^{-1}(\eta_{ps}) = \frac{\exp(\theta_p + b_s)}{1 + \exp(\theta_p + b_s)}$ (2)
Log-normal model	
$E(t_{ps}\|\tau_p, \delta_s) = \delta_s - \tau_p + \varepsilon$ (3)	$E(t_{ps}\|\tau_p, \delta_s) = \delta_s + \tau_p + \varepsilon$ (4)

Note: μ_{ps}: Probability of a correct response for respondent p on stimulus s given the linear combination of predictors η_{ps}; η_{ps}: Linear combination of predictors $\theta_p + b_s$; $logit^{-1}$: Inverse of the *logit* link function $logit = log\left(\frac{\mu_{ps}}{1-\mu_{ps}}\right)$; θ_p: Person parameter of the Rasch model, b_s: Stimulus parameter of the Rasch model, τ_p: Person parameter of the log-normal model, δ_s: Time intensity parameter of the log-normal model. The model in Equation 2 is a generalized linear model.

the higher the amount of time required by stimulus s to get a response. In a LM, the expected value of the dependent variable (in this case, the log-time response) of respondent p to stimulus s is directly linked to the linear combination of predictors η_{ps} by an *identity* function. The log-normal model (Equation 3 in Table 1) can be equated to a LM (Equation 4 in Table 1).

In the typical formulation of Rasch and log-normal models, respondent and stimulus parameters move in opposite directions (i.e., $\theta_p - b_s$ and $\delta - \tau_p$ in Equations 1 and 3, respectively). However, when respondent and stimulus parameters are estimated by using (G)LMs, their estimates move in the same direction, hence resulting in an additive effect (i.e., $\theta_p + b_s$ in Equation 2 and $\delta_s + \tau_p$ in Equation 4, see Table 1). The item parameter b_s can no longer be interpreted as an impediment property of the stimulus (difficulty) but it should be interpreted as a facilitation property of the stimulus (easiness) (De Boeck et al., 2011; Doran et al., 2007). This is consistent with the easiness parametrization of linear test models (see e.g., Fischer, 1973). The speed parameter τ_p has a reverse interpretation as well, such that the lower the value of parameter τ_p,

the faster respondent p is.

Random Factors and Random Effects in Linear Mixed-Effects Models

When there are reasons to believe that there are sources of variability generating dependencies between the observations, such as in data with fully-crossed structures, random effects addressing the uncontrolled random variability at different levels should be included in the linear component $\eta_{ps} = \mathbf{X}\beta$, where $\mathbf{X}$ represents the model matrix of the fixed factors (intercepts and slopes) and β indicates their coefficients (i.e., fixed effects). The specification of the random structure makes it possible to decompose the error variance into different levels, which correspond to the levels where the uncontrolled sources of random variation can be reasonably found (Doran et al., 2007). These levels can be understood as the random effects of the model, which reflect the assumptions on the structure of dependency between the observations and allow for acknowledging the multilevel structure of the data (Barr et al., 2013). To include random factors (e.g., stimuli and respondents) in a linear model, that is, to switch from a linear model to a linear mixed-effects model, the linear component needs to be extended from $\eta_{ps} = \mathbf{X}\beta$ to $\eta = \mathbf{X}\beta + \mathbf{Z}d$, where $\mathbf{Z}$ is the $P \times Q$ matrix of the random effects (i.e., Q is the dimension of the random factors vector), and d is the vector of the random effects. The dimension Q of d derives from the number of levels of each random factor and their combinations. For instance, if respondents are considered as random factors, the dimension Q of the random effects vector will have as many levels as the number of respondents. Consequently, the dimension of d can be potentially very large (Doran et al., 2007). 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 $\mathbf{\Sigma}$, which is determined by a single vector of parameters Γ (Doran et al., 2007). The dimension of Γ is usually rather small, and its size derives from the number of random factors specified in the model, regardless of the number of levels they include. For instance, consider a model in which overall (i.e., across conditions) respondent variability, overall item variability, and item variability in three different conditions are accounted for. In this model, five random factors are considered and their related random effects are specified. In particular, one random effect for overall respondent variability, one for overall item variability, and one allowing for

the multidimensionality of item variability in the three conditions are specified. The dimension of the vector parameter Γ for this model is 5, and it remains 5 regardless of the number of respondents or items used. If the fixed effects of the experimental conditions are added to the model matrix, then β will contain 3 parameters (i.e., the fixed intercept and the slopes that describe the deviation of each of the other two conditions from the fixed intercept).

Random Structures

The objective of LMMs is to estimate the parameters of the fixed effects as defined in vector β and the parameters of the random effects as defined in vector Γ. Consequently, the parameters estimated for the random factors are not the parameters associated to each level of each factor, but the variance-covariance matrices of the populations from which the random factors are drawn. A measure for each level of the random factor in d is obtained in the form of *conditional modes*, which are the values maximizing the conditional density of the random effects given the vectors of fixed and random parameters and the observed data (Doran et al., 2007). The conditional modes that describe the deviation of each random factor from the fixed factors are not properly parameters of the models (see, e.g., Pastore, 2015) and are usually referred to as *Best Linear Unbiased Predictors* (BLUPs, Doran et al., 2007; Pinheiro & Bates, 2006). In the application presented in this tutorial, BLUPs are used in conjunction with the estimates of the fixed effects to obtain the respondent and stimulus estimates of Rasch-like and log-normal models. As such, Rasch and log-normal parametrizations depend on the random structure of the model, which in turn depends on the variability observed in the data.

The total number of parameters of the model is determined by the dimension of β (i.e., the coefficients associated to the model matrix of the fixed effects) and that of Γ (i.e., the vector of parameters of the random effects). As such, if entropy indexes like the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978) are used for model comparison, the penalization of the log-likelihood will depend on the dimension of β and on that of Γ. The number of levels of the random effects, whose deviations from the estimates of the fixed effects are expressed by the BLUPs, does not affect the penalization of the models. Considering the example presented in the previous section, the dimension of β was

3 and that of Γ was 5. As such, the total number of parameters of the model was 8, regardless of the actual number of respondents or items. The complexity of the model in terms of number of estimated parameters remains unvaried as long as no fixed effects are added to the model matrix and the random structure is left unaltered.

Potentially, (G)LMMs allow for accounting for all the possible sources of variability that are expected in data with fully-crossed structures (i.e., Maximal Models, Barr et al., 2013). However, such complex models are at risk of convergence failure because the random structure is too complex to be supported by the data (Bates, Kliegl, et al., 2015). As such, only the random structures that are meaningful for obtaining Rasch-like parametrizations of response accuracies and times of fully-crossed data are presented in this tutorial. The estimates of person parameters (θ_p and τ_p) derive from the random effects of the respondents, being either $\alpha_p \sim \mathcal{N}(0, \sigma^2_{\alpha_p})$ (random intercepts, Models 1 and 2 in Table 2, where $\mathcal{N}$ identifies a normal distribution) or $\beta_{pc} \sim \mathcal{MVN}(0, \Sigma_{pc})$ (random slopes in the contrasting conditions c, Model 3 in Table 2, where $\mathcal{MVN}$ identifies a multivariate normal distribution). The estimates of stimulus parameters (b_s and δ_s) derive from the random effects of the stimuli, being either $\alpha_s \sim \mathcal{N}(0, \sigma^2_{\alpha_s})$ (random intercepts, Models 1 and 3 in Table 2) or $\beta_{sc} \sim \mathcal{MVN}(0, \Sigma_{sc})$ (random slopes in the contrasting conditions c, Model 2 in Table 2). In the LMMs, the distribution of the error term follows a normal distribution, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. In the GLMMs, the probability of observing a correct response is conditional to the parameters of the respondents and of the stimuli, following a logistic distribution. Consequently, only the random structures of the LMMs include the error term ε. Table 2 summarizes the Rasch-like and log-normal parameters attainable from each model random structure, along with the `R` code for their estimation.

Table 2

Overview of the Estimable Parameters and `lme4` Notation.

Model	Formal notation	`lme4` notation	p parameters	s parameters
1	$\alpha + \beta_c X_c + \alpha_{p[i]} + \alpha_{s[i]}$	`y ~ 0 + condition + (1|stimuli) + (1|respondents)`	θ_p or τ_p	b_s or δ_s
2	$\alpha + \beta_c X_c + \alpha_{p[i]} + \beta_{s[i]} c_i$	`y ~ 0 + condition + (0 + condition|stimuli) + (1|respondents)`	θ_p or τ_p	b_{sc} or δ_{sc}
3	$\alpha + \beta_c X_c + \alpha_{s[i]} + \beta_{p[i]} c_i$	`y ~ 0 + condition + (1stimuli) + (0 + condition|respondents)`	θ_{pc} or τ_{pc}	b_s or δ_s

Note: Respondent $p = 1, \ldots, P$, Stimulus $s = 1, \ldots, S$, Condition $c = 1, \ldots, C$, where P, S, and C, denote the number of respondents, stimuli, and conditions, respectively. θ: respondent ability estimates, τ: respondent speed estimates, b: stimulus easiness estimates, δ: stimulus time intensity estimates. The dependent variable y can be either the response accuracies in the GLMMs or the log-time responses in the LMMs. The fixed intercept α is set at 0. The error term ε_i, where i denotes the single trial, must be added to the notation when LMMs are considered. The notation `0 + condition` is used to set the fixed intercept $\alpha = 0$, such that the marginal log-odds of the probability of responding correctly in the two conditions or the marginal means of the response times in the two conditions are estimated in GLMMs and LMMs, respectively. The notations `(1|stimuli)` and `(1|respondents)` are used to specify the random intercepts of the stimuli and of respondents, respectively. The notations `(0 + condition|stimuli)` and `(0 + condition|respondents)` are used to specify the random slopes of the stimuli and the respondents in the conditions, respectively. The notations `(0 + condition||stimuli)` or `(0 + condition||respondents)` can be used to force the independent estimation of the variances of either the stimuli or the respondents in the two conditions, respectively. The meaning and interpretation of the parameters are illustrated in the Section titled "Random Structures". The `R` code is fully explained in the Section titled "`R` Code for Fitting the (G)LMMs".

The `R` code is fully commented and detailed in Section "`R` Code for Fitting the (G)LMMs", while in this section the focus is in illustrating the bridge between the mathematical notation of the models and the corresponding specification in `R`. The fixed structure (i.e., the fixed effect of the contrasting conditions) is equal across models. Specifically, the fixed intercept α has been set to 0 (i.e., `0 + condition`, where `condition` is the variable with two levels that identifies the contrasting conditions), so that the marginal log-odds of the probability of responding correctly in the two conditions (GLMMs) and the marginal mean of the log-response times in the two conditions (LMMs) are estimated.

Model 1 is the null model, where only the random intercepts of both the respondents (i.e., `(1|respondents)`) and the stimuli (i.e., `(1|stimuli)`) are specified along with the fixed effect of the contrasting conditions (i.e., `0 + condition`). This model should be preferred when low within–respondent between–condition variability and low within–stimulus between–condition variability are observed. The lack of variability at both levels might indicate a lack of the effect of the contrasting conditions on both the performance of the respondents and the functioning of the stimuli. Overall (i.e., across–conditions) respondent and stimulus estimates of Rasch-like and log-normal models are obtained by considering the BLUPs that describe the deviation of each respondent and each stimulus from the estimate of the fixed intercept. Since in this application the fixed intercept is set to 0, the BLUPs describe the deviations from 0. The overall respondent estimates are a measure of the individual differences of the respondents and they can be used for further analyses, while the overall stimulus estimates can be used for the investigation of the functioning of the stimuli related to their properties and characteristics. Specifically, in experiments such as the IAT or those for evaluating the SNARC effect, the stimuli are supposed to be prototypical of the category to which they belong (i.e., they are supposed to be easily recognized as representative of that category). In this sense, the stimuli belonging to the same category should show similar estimates. By comparing the estimates of the stimuli belonging to the same category, it should be possible to investigate whether they are all easily recognizable as prototypical exemplars or not.

In Model 2, the random slopes of the stimuli in the contrasting conditions (i.e., `(0 + condition|stimuli)`) are specified along with the random intercepts of the respondents

across contrasting conditions (i.e., `(1|respondent)`) to account for the within–stimulus between–condition variability and for the overall between–respondent variability. This model is expected to be the best fitting one when high within–stimuli between–conditions variability is observed and respondents show a low between–conditions variability. This might suggest that the variations observed between the contrasting conditions are mostly ascribable to the variability in the functioning of the stimuli. Overall respondent estimates and condition–specific stimulus estimates of Rasch-like and log-normal models are obtained by considering the BLUPs of the respondents and the stimuli. The BLUPs of the respondents describe the deviation of each respondent from the fixed intercept. The BLUPs of the stimuli describe the deviation of each stimulus from the estimates of the fixed effects of the conditions, which are either the marginal log-odds of the probability of responding correctly (GLMM) or the mean response times (LMM) in each condition. In this case, the overall respondent estimates can be used for the investigation of the individual differences, regardless of the effect of the contrasting conditions. The high within–stimulus between–condition variability that allows for the estimation of condition–specific stimulus parameters suggests that the functioning of the stimuli vary between the conditions. The differential measure computed on the condition–specific stimulus estimates can inform about the contribution of each stimulus to the effect under investigation. For instance, in the IAT case, this might lead to a better understanding of the automatic associations driving the effect (e.g., Anselmi et al., 2011; Epifania et al., 2022a).

The random structure of Model 3 has the same degree of complexity as the random structure of Model 2, but the multidimensionality on the error term is specified at the respondent level. Model 3 accounts for the within–respondent between–condition variability and the overall between–stimulus variability by specifying the random slopes of the respondents in the contrasting conditions (i.e., `(0 + condition|respondents)`) and the random intercepts of the stimuli across the contrasting conditions (i.e., `(1|stimuli)`). If Model 3 results as the best fitting one, it might be speculated that the variations observed between the contrasting conditions are mostly ascribable to the variability in the performance of the respondents. Overall stimulus estimates and condition–specific respondent estimates of Rasch-like and log-normal models are obtained. The overall stimulus estimates are obtained from the BLUPs that

describe the deviation of each stimulus from the estimate of the fixed intercept. The condition–specific estimates of the respondent are obtained by considering together the deviation of each respondent from the fixed estimates of each condition (i.e., BLUPs) and the fixed estimates of each condition. In this case, the overall stimulus estimates can be used for the investigation of the functioning of the stimuli, regardless of the effect of the contrasting conditions. The high within–respondent between–condition variability suggests that the performance of the respondents is affected by the contrasting conditions. A measure of the bias on the performance due to the contrasting conditions can be obtained by computing the difference between the condition–specific estimates of the respondents.

Estimating LMMs With `R`

In this section, a brief description of the data format needed for fitting the LMMs is given. Then, schematic overviews of the `R` codes to be used for estimating the (G)LMMs presented in the previous section and for obtaining Rasch-like and log-normal parametrizations are provided.

Wide Format and Long Format

Tables 3 and 4 present data in wide and long formats, respectively, to help the reader visualize the differences between the two formats. These tables contain the same information, namely the accuracy and response time of the responses given by three respondents (p_1, p_2, p_3) to two stimuli (☺, 😜) in two different conditions ("Condition A" and "Condition B").

When data are organized in a $P \times S$ matrix (i.e, wide format, Table 3), each row represents a respondent $p = 1, \ldots, P$ and each column represents a stimulus $s = 1, \ldots, S$. Each cell contains the response of each respondent to a stimulus. In Table 3, the three rows represent the three respondents, whereas the columns contain the accuracies and response times of the responses given by each respondent to stimulus ☺ and stimulus 😜 in the two contrasting conditions ("Condition A" and Condition B").

The same information presented in Table 3 is reported in Table 4 but it is organized in a long format. For each respondent, there are as many rows as number of trials that they have been presented with, such that the total number of rows is obtained as the product between

Table 3
Data in Wide Format.

	Condition A				Condition B			
p	RT ☺	Accuracy ☺	RT 🤪	Accuracy 🤪	☺	Accuracy ☺	RT 🤪	Accuracy 🤪
p_1	520	1	320	1	420	0	500	1
p_2	320	0	400	0	620	0	380	1
p_3	720	1	600	1	520	1	570	0

Note: p: Respondents, RT: Response times in milliseconds, Accuracy: Accuracy of the responses where 1 and 0 denote correct and incorrect responses, respectively.

the number of respondents and the number of trials in both conditions. For instance, in the example in Table 3, each respondent has been presented with each stimulus (☺, 🤪) twice, one in Condition A and one in Condition B. As such, there are four observations for each respondent represented by the four rows for each respondent in Table 4. Response times and accuracies pertaining to each ☺ and each 🤪 in the two conditions are recorded in their respective columns, namely "RT" and "Accuracy".

Table 4
Data in Long Format.

Respondent	Condition	Stimulus	RT	Accuracy
p_1	A	☺	520	1
p_1	B	☺	420	0
p_1	A	🤪	320	1
p_1	B	🤪	500	1
p_2	A	☺	320	0
p_2	B	☺	620	0
p_2	A	🤪	400	0
p_2	B	🤪	380	1
p_3	A	☺	720	1
p_3	B	☺	520	1
p_3	A	🤪	600	0
p_3	B	🤪	570	0

Note: p: Respondents, RT: Response times in milliseconds, Accuracy: Accuracy of the responses where 1 and 0 denote correct and incorrect responses, respectively.

The data set for the application of the `R` codes in the next section must be in long for-

mat. Usually, the majority of software used for the administration of tasks in experimental psychology (e.g., PsychoPy, Inquisit, E-Prime) provides data in long format. However, if the data are not in long format, `R` allows for reshaping the data format from wide to long (e.g., the `reshape()` function from the `stats` package).

`R` Code for Fitting the (G)LMMs

The code presented in this section can be executed without changes as long as the data set on which the models are applied contains the following variables stored in columns with the following labels:

- `respondent`: Column containing respondent IDs (the ID can be numeric, factor, or string, as long as it is unique for each respondent).
- `condition`: Column containing the labels of the contrasting conditions (factor with two levels, such as `ConditionA` and `ConditionB`).
- `stimuli`: Column containing the labels identifying each stimulus (e.g., `stimulus1`, `stimulus2`, `stimulus3`, `stimulus4`).
- `latency`: Column containing the response times to each trial. Response times can be expressed in any time scale. As it will be shown in the application examples, knowing the time scale is fundamental for interpreting the log-normal model estimates.
- `correct`: Column containing the accuracy of the response to each trial, where 0 denotes an incorrect response and 1 denotes a correct response.

If the column labels are renamed, then the following code must be changed accordingly.

The (G)LMMs are estimated with the `lme4` package (Bates, Machler, et al., 2015). Once installed with the code `install.packages("lme4")`, the code `library(lme4)` must be run to make the package available for use.

Table 5 summarizes the main code differences for the application of the GLMMs and of the LMMs.

Table 5
`lme4` Code for the Application of (Generalized) Linear Mixed Effects Models on Accuracies and Response Times.

Accuracy	Response Time
GLMM	LMM
`correct`	`log(latency)`
`glmer()`	`lmer()`
`family = "binomial"`	`REML = FALSE`

Note: GLMM: Generalized linear mixed-effects model; LMM: Linear mixed-effects model; `correct`: Response accuracy in the data set in long format; `log(latency)`: Logarithm of the response times in the data set in long format; `glmer()`: Function for estimating the GLMMs in the `lme4` package; `lmer()`: Function for estimating the LMMs in the `lme4` package.

The argument `family = "binomial"` of the `glmer()` function for estimating GLMM from the response accuracies specifies the link function. In this case, the default for the binomial family is the logit link function. The argument `REML = FALSE` of the `lmer()` function for estimating LMM from the response times produces maximum likelihood estimates of the model parameters.

The fixed and random effects of the models and the `lme4` notation for the specification of their random structures have been illustrated in Table 2. The combination of the `lme4` notation in Table 2 with the specific arguments in Table 5 results in the complete code for estimating all the models. While the full code for the estimation of these models on two real data sets is made available as supplementary materials, some examples are illustrated in what follows.

For instance, Model 1 (Table 2) can be estimated from the accuracies as:

```
m_accuracy1 <- glmer(correct ~ 0 + condition + (1|stimuli) + (1|respondent),
        data = data,
        family = "binomial")
```

and from the log-transformed response times as:

```
m_time1 <- lmer(log(latency) ~ 0 + condition + (1|stimuli) + (1|respondent),
        data = data,
        REML = FALSE)
```

As in any linear model syntax in R, the linear combination of the predictors is on the right side of the tilde while the dependent variables are on the left side. To set the fixed intercept at 0 and obtain the marginal log-odds of the probability of responding correctly or the marginal mean of the log-latency in the contrasting conditions, the code `0 + condition` is used. The random intercepts of the respondents and the stimuli can be specified as `(1|respondent)` and `(1|stimuli)`, respectively. The function `log()` is used to apply the logarithm transformation on the response times. The data from which the models are estimated are specified with `data = data`, and it must be changed according to the name of the object into which the data are stored. The three models presented in Table 2 can be estimated by using the R code to define the specific random structures of the GLMMs and LMMs. For instance, Model 2 (Table 2) on the response accuracies can be estimated as:

```
m_accuracy2 <- glmer(correct ~ 0 + condition + (0 + condition|stimuli) + (1|respondent),
        data = data,
        family = "binomial")
```

and Model 3 from Table 2 on the log-time responses can be estimated as:

```
m_time3 <- lmer(log(latency) ~ 0 + condition + (1|stimuli) + (0 + condition|respondent),
        data = data,
        REML = FALSE)
```

In `m_accuracy2` and `m_time3`, the random intercepts of the respondents and those of the stimuli are specified as in `m_accuracy1` or `m_time1`. The random slopes of the stimuli in `m_accuracy2` are specified as `(0 + condition|stimuli)`, while the random slopes of the respondents in `m_time3` are specified as `(0 + condition|respondent)`.

Model comparison allows for choosing the model with the best fitting random structure, from which Rasch-like and log-normal parametrizations are obtained. Since Models 2 and 3 have the same degrees of freedom, the χ^2 statistics obtained from their comparison with the function `anova()` cannot be used for choosing the best fitting model, and comparative fit indexes should be used instead. In this tutorial, the `anova()` function is used for the convenience of having comparative fit indexes, deviance, log-likelihood and degrees of freedom of the models displayed together.

Given that the models on accuracies and response times are estimated separately, the best fitting model on response accuracies can be different from the best fitting model on log-time responses. This might suggest that the performance expressed by either the accuracies or the response times is affected by the independent variable, while the other is not. In other words, the processes involved in the accuracy performance might be different from those involved in the time performance. For instance, it is plausible that a respondent would slow down in the condition that is against their automatic association in order to keep the accuracy performance unaltered between the two conditions, or vice-versa (i.e., they might alter their accuracy performance to keep their speed unaltered between conditions).

R Code for Obtaining Rasch-Like and Log-Normal Parametrizations

This section presents the code for extracting person and stimulus estimates according to the random structure of the best fitting models. Since the code is identical regardless of the starting model (i.e., GLMM or LMM), the generic object `model#` is used for illustration purposes, but it should be replaced with the specific name of the model (e.g., `m1_accuracy`, `m1_time`). Rasch-like and log-normal parametrizations of the data are obtained with the `coef()` function, which adds up together the BLUPs of each respondent and each stimulus to the estimates of the fixed effects.

If Model 1 is the best fitting model, only overall (i.e., across contrasting conditions) respondent (θ_p or τ_p) and stimulus (b_s or δ_s) estimates can be obtained from the fitted model. Respondent estimates are obtained as:

```
estimates_p_m1 <- data.frame(
        respondent = rownames(coef(model_1)$respondent),
        estimate = coef(model_1)$respondent[, 1]
        )
```

and stimulus estimates are obtained as:

```
estimates_s_m1 <- data.frame(
        stimuli = rownames(coef(model_1)$stimuli),
         estimate = coef(model_1)$stimuli[, 1]
```

```
)
```

The data frames `estimates_p_m1` and `estimates_s_m1` contain respondent and stimulus estimates arranged in wide formats, such that the number of rows in `estimates_p_m1` is equal to the number of respondents, and the number of rows in `estimates_s_m1` is equal to the number of stimuli. Columns `respondent = rownames(coef(model_1)$respondent)` and `stimuli = rownames(coef(model_1)$stimuli)` contain the identifies of the respondents and the stimuli, respectively. Column `estimate` contains either the respondent estimates (θ_p or τ_p) or the stimulus estimates (b_s or δ_s). The syntax `[,1]` selects the intercepts corresponding to the BLUPs of the respondents or the stimuli from the data frame of the random effects.

If Model 2 is the best fitting one, high variability in the functioning of the stimuli between the two contrasting conditions is expected. Overall (i.e., across conditions) respondent estimates (i.e., θ_p or τ_p) and condition–specific stimulus estimates (i.e., b_{sc} and δ_{sc}) are obtained. The same code as in Model 1 can be used for obtaining the respondent estimates, while the condition–specific stimulus estimates are obtained with:

```
estimates_s_m2 <- coef(model_2)$stimuli[, -1]
```

The number of rows in `estimates_s_m2` is equal to the number of stimuli and the number of columns is equal to the total number of experimental conditions. The command `[,-1]` removes the first column of the matrix (i.e., the fixed intercept set to 0).

Finally, if Model 3 is the best fitting model, high variability in the performance of the respondents between conditions is expected. In this case, overall (i.e., across conditions) stimulus estimates (i.e., b_s or δ_s) and condition–specific respondent estimates (i.e., θ_{pc} or τ_{pc}) are obtained. The same code as in Model 1 can be used for obtaining the stimulus estimates, while the condition–specific respondent estimates are obtained with:

```
estimates_p_m3  <- coef(model_3)$subject[, -1]
```

where `estimates_p_m3` has a number of rows equal to the number of respondents, a number of columns equal to the total number of conditions.

Examples of Application

This section presents application examples of the modeling approach on data sets from two experiments pertaining to: (i) the investigation of a SNARC-like effect focused on the spatial perception of weights (Vicovaro & Dalmaso, 2021), and (ii) the implicit investigation of racial prejudice among university students with the IAT (Epifania et al., 2020b, 2021). In the former example, the focus is on the SNARC-like effect considered at the sample level, while in the latter one the focus is on the individual differences of the respondents. This approach allows for obtaining estimates at the sample level (i.e., the fixed effects) while controlling for the variability at respondent and stimulus levels, hence overcoming the shortcomings associated with the by-participant approach often used for analyzing these data (see, e.g., Barr et al., 2013; Judd et al., 2012, 2017). The compatible and incompatible conditions are defined a priori at the sample level, where the compatible condition is the one consistent with the SNARC effect. In the latter example, the focus is on the individual differences of the respondents, and this approach allows for obtaining detailed information on the effect of the contrasting conditions at the sample level and at the level of the single respondent, while acknowledging and exploiting the variability at the stimulus level. Contrary to the previous example, the definition of compatible and incompatible conditions may vary between respondents. The commented `R` script is available as supplementary material. All the graphical representations are obtained with the `ggplot2` package (Wickham, 2016).

SNARC-Like Effect

Data are published in Vicovaro and Dalmaso (2021). The authors investigated a SNARC-like effect entailing the perception of weights, under the assumption that heavier weights are perceived at the bottom and lighter weights are perceived at the top of a vertical axis. The investigation pertained several experiments, and the data used for this application refers to Experiment 2.

The first rows of the data set (Figure 2) can be obtained with the `head()` function. In this experiment, respondents were presented with different stimuli (column `stimulus` of the data set in the figure) that could be either light or heavy (column `weight` in the figure). The

stimuli appeared one at the time on the computer screen, and the respondents had to recognize and classify them as light or heavy weights via an ad hoc vertical apparatus with a response key on its top and on its bottom. The response keys associated to light and heavy weights defined the contrasting conditions of the experiment (column `condition` in the figure). In the compatible condition, the top response key was used to sort the light weights, while the bottom key was used to sort the heavy weights. The response keys associated to light and heavy weights were reversed in the incompatible condition. The response times in milliseconds and correctness (1: correct, 0: incorrect) of each trial were registered (column `time` and `correct` in the figure, respectively).

```
> head(data)
  subject weight stimulus  condition resp time correct
1       1  light     1 kg Compatible    2  726       1
2       1  heavy      3 t Compatible    3  535       1
3       1  heavy     8 kg Compatible    3  520       1
4       1  light     2 kg Compatible    2  917       1
5       1  light      9 g Compatible    2  596       1
6       1  light      6 g Compatible    2  525       1
```

Figure 2: Head of the SNARC-like data set

The data set contains the observations on 32 respondents. Sixteen stimuli were used (eight light weights, eight heavy weights), each of which was presented fourteen times in each condition. As such, each respondent was presented with 224 trials in each condition, for a total of 448 trials across conditions. Further details on the experiment can be found in Vicovaro and Dalmaso (2021).

The models can be estimated from the accuracies and response times of this data, and they can be compared with the `anova()` function. The comparisons between the accuracy models and those between the log-time models are illustrated in top and bottom panels of Figure 3, respectively.

The first five lines of both panels report the name of the data set from which the models were fitted (`Data: data`) and the code of the models (`Models:`). The table of model comparison follows, where the rows contain the labels of each model and the columns report the information of interest, namely the number of estimated parameters for each model (`npar`), the AIC (`AIC`) and the BIC (`BIC`), the Log-Likelihood (`logLik`), the Deviance (`deviance`), and the χ^2 with its associated degrees of freedom, test, and p-value (`Chisq`, `df`, `P(>Chisq)`,

```
> anova(m_accuracy1, m_accuracy2, m_accuracy3)
Data: data
Models:
m_accuracy1: correct ~ 0 + condition + (1 | subject) + (1 | stimulus)
m_accuracy2: correct ~ 0 + condition + (1 | subject) + (0 + condition | stimulus)
m_accuracy3: correct ~ 0 + condition + (0 + condition | subject) + (1 | stimulus)
            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
m_accuracy1    4 3510.4 3540.6 -1751.2   3502.4
m_accuracy2    6 3505.9 3551.3 -1747.0   3493.9  8.440  2     0.0147 *
m_accuracy3    6 3483.2 3528.6 -1735.6   3471.2 22.678  0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(m_time1, m_time2, m_time3)
Data: data
Models:
m_time1: log(time) ~ 0 + condition + (1 | subject) + (1 | stimulus)
m_time2: log(time) ~ 0 + condition + (1 | subject) + (0 + condition | stimulus)
m_time3: log(time) ~ 0 + condition + (0 + condition | subject) + (1 | stimulus)
        npar     AIC     BIC logLik deviance    Chisq Df Pr(>Chisq)
m_time1    5 -3425.4 -3387.6 1717.7  -3435.4
m_time2    7 -3423.7 -3370.8 1718.9  -3437.7   2.3064  2     0.3156
m_time3    7 -3879.8 -3826.8 1946.9  -3893.8 456.0634  0
```

Figure 3: Model comparison of the accuracy models (top panel) and the log-time models (bottom panel) applied to the SNARC-like data.

respectively). Generally speaking, the best fitting model is the one with the lowest BIC, AIC, deviance, and highest log-likelihood. Among accuracies and response times models and log-time models, Models 3 present the lowest AIC, BIC, deviance, and the highest log-likelihood. As such, Model 3 is chosen as the best fitting one for both accuracies (`m_accuracy3`) and log-time (`m_time3`) responses. Since Model 3 results as the best fitting one on both, it might be speculated that the variability in the responses between the contrasting conditions is mostly due to the variability in the accuracy and time performance of the respondents.

The summary of the best fitting models can be obtained with the `summary()` function. The summary of `m_accuracy3` and that of `m_time3` are depicted in Figures 4 and Figure 5, respectively.

The first sections of the outputs in Figures 4 and 5 inform about the method used for fitting the models. In Figure 4, information on the link function is given as well (`Family: binomial ( logit )`). The formula of the fitted model, the comparative fit indexes, and the descriptive statistics of the residuals follow.

The estimates of the fixed effects are interpretable as the estimates of a linear model, bearing in mind that the estimates associated to each level of the predictor (i.e., the contrasting

```
> summary(m_accuracy3)
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: correct ~ 0 + condition + (0 + condition | subject) + (1 | stimulus)
   Data: data

     AIC      BIC   logLik deviance df.resid
  3483.2   3528.6  -1735.6   3471.2    14260

Scaled residuals:
     Min       1Q   Median       3Q      Max
-15.3177   0.0864   0.1207   0.1800   0.9829

Random effects:
 Groups   Name                   Variance Std.Dev. Corr
 subject  conditionCompatible    1.3480   1.1610
          conditionIncompatible  0.7166   0.8466   0.83
 stimulus (Intercept)            0.3860   0.6213
Number of obs: 14266, groups:  subject, 32; stimulus, 16

Fixed effects:
                      Estimate Std. Error z value Pr(>|z|)
conditionCompatible     4.0003     0.2759   14.50   <2e-16 ***
conditionIncompatible   4.2026     0.2405   17.48   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            cndtnC
cndtnIncmpt 0.761
```

Figure 4: Summary of the results of Model 3 for the response accuracies to the SNARC-like data.

condition) are interpretable as the marginal log-odds of a success in each condition (Figure 4) or as the marginal mean of the log-time response in each condition (Figure 5). Figure 4 reports the significance of the fixed effects. Both estimates are significantly different from 0, indicating that high proportions of correct responses are expected in both conditions. Moreover, the estimates are close to each other. The log-odds can be transformed in probability values with the inverse of the *logit* link function, $\mu_{ps} = \frac{\exp(\eta_{ps})}{1+\exp(\eta_{ps})}$. The probabilities of responding correctly in the two conditions are .98 and .99. Differently from Figure 4, Figure 5 does not report the significance of the estimates. Indeed, the significance of the fixed effects in the GLMM is tested with a z-test based on the standardized normal distribution. As such, a critical z-value can be obtained given the Type I Error probability α, and the computation of the p-value is straightforward. Conversely, within a LMM, a t-test based on the Student's t probability distribution is employed to evaluate the significance of the fixed effects, which requires the determination of

```
> summary(m_time3)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(time) ~ 0 + condition + (0 + condition | subject) + (1 |      stimulus)
   Data: data

     AIC      BIC   logLik deviance df.resid
 -3879.8  -3826.8   1946.9  -3893.8    14259

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.5492 -0.6234 -0.1034  0.4549  6.2299

Random effects:
 Groups   Name                   Variance Std.Dev. Corr
 subject  conditionCompatible    0.011608 0.10774
          conditionIncompatible  0.016220 0.12736  0.76
 stimulus (Intercept)            0.002957 0.05437
 Residual                        0.043612 0.20883
Number of obs: 14266, groups:  subject, 32; stimulus, 16

Fixed effects:
                     Estimate Std. Error t value
conditionCompatible    6.52582    0.02353   277.3
conditionIncompatible  6.48035    0.02642   245.3

Correlation of Fixed Effects:
            cndtnC
cndtnIncmpt 0.825
```

Figure 5: Summary of the results of Model 3 for the log-time responses to the SNARC-like data.

a critical t value for the computation of the p-value. The t-value is determined by considering both the Type I Error probability α and the degrees of freedom. However, obtaining accurate degrees of freedom for data with a fully-crossed structure can be challenging, impeding the straightforward identification of the appropriate critical t-value (Bates, Machler, et al., 2015). Consequently, conducting a significance test may not be feasible, leading to the omission of p-value reporting in the output (Bates, Machler, et al., 2015). Nevertheless, the `lmerTest` package (Kuznetsova et al., 2017) offers an approximation for computing degrees of freedom, thereby enabling the significance testing of model estimates.

The summary tables of the random effects report the estimates of variances (Column `Variance`) and standard deviations (Columns `Std.Dev`) of the populations from which the random factors are supposed to be drawn. Column `Groups` identifies the random factors, while column `Name` specifies the type of random factor (either random slopes or random intercepts). Column `Corr` refers to the covariation between the slopes in the random factor. According to Figure 4, the responses presented more variability in the compatible condition (1.3480) than

in the contrasting one (0.7166), and the correlation between the conditions was high (0.83). Also the stimuli presented a good variability (0.3860). Concerning the log-time model, Figure 5 shows that the variability of the respondents tended to be particularly low in both conditions (0.011608 and 0.016220 in the compatible and incompatible conditions, respectively), as well as the stimulus variability across conditions (0.002957). As in the accuracy model, the correlation between the conditions was high (0.76). The ability and speed estimates are obtained from the respective models as:

```
# ability estimates from the accuracy model
ability_m3  <- coef(m_accuracy3)$subject[, -1]
# speed estimates from the log-time model
speed_m3  <- coef(m_time3)$subject[, -1]
```

while the easiness and time intensity estimates are obtained as:

```
# easiness estimates
easiness_m3 <- data.frame(
        stimuli = rownames(coef(m_accuracy3)$stimuli),
        estimate = coef(m_accuracy3)$stimuli[, 1]
        )
# time intensity estimates
intensity_m3 <- data.frame(
        stimuli = rownames(coef(m_time3)$stimuli),
        estimate = coef(m_time3)$stimuli[, 1]
        )
```

The Rasch-like and log-normal model estimates are illustrated in the top and bottom panels Figure 6, respectively. These graphical representations can be obtained from the R objects containing the estimates of respondents and stimuli, but the data frame containing the condition–specific estimates of the respondents must be reshaped in long format. For instance, the code for reshaping the data frame containing the ability estimates (and for changing the column names) is as follows:

```
# reshape the data frame with ability estimates from wide to long
```

```
ability_long = stack(ability_m3)

# rename the columns

colnames(ability_long) = c("theta", "condition")
```

The code for further customizing the graphical representations illustrated in this section is reported in the script provided in the supplementary material. The basic code for obtaining the graphical representation in Figure 6a is:

```
ggplot(ability_long, # the data frame with the ability estimates in long format
       aes(x = theta, fill = condition)) +
       geom_density()
```

The estimates of θ_p are reported on the x-axis, and the density function is determined by the `geom_density()` function. Since the `fill = condition` argument is specified, the density of θ is determined according to the conditions. The lower the value of θ (x-axis in Figure 6a), the lower the ability (i.e., the lower the number of correct responses provided by the respondents). Since the density distributions of the condition–specific ability parameters in Figure 6a almost overlap, the ability of the respondents appeared to be similar in both conditions. Nonetheless, the distribution in one of conditions shows a bi-modal distribution, suggesting that respondents tended to settle around two levels of performance, one better than the other. This effect is not observed in the other condition.

The plot in Figure 6b is obtained from the data frame that contains the easiness estimates `easiness_m3` with the code:

```
ggplot(easiness_m3, aes(x = easiness, y = stimuli)) + geom_point()
```

The estimates b_s of the stimuli are reported on the x-axis, while the labels of the stimuli are on the y-axis. The command `geom_point()` draws a point to represent the easiness estimate of each stimulus. The higher the value b of stimulus s, the less difficult the stimulus (i.e., the higher the number of correct responses registered on s). In this case, the variability of the easiness estimates ranges between -1 and $+1$, although majority of the estimates are close or above the mean (i.e., 0).

The plot of the speed estimates τ of the respondents (Figure 6c) can be obtained by appropriately replacing the objects and their labels in the code used for plotting the ability estimates

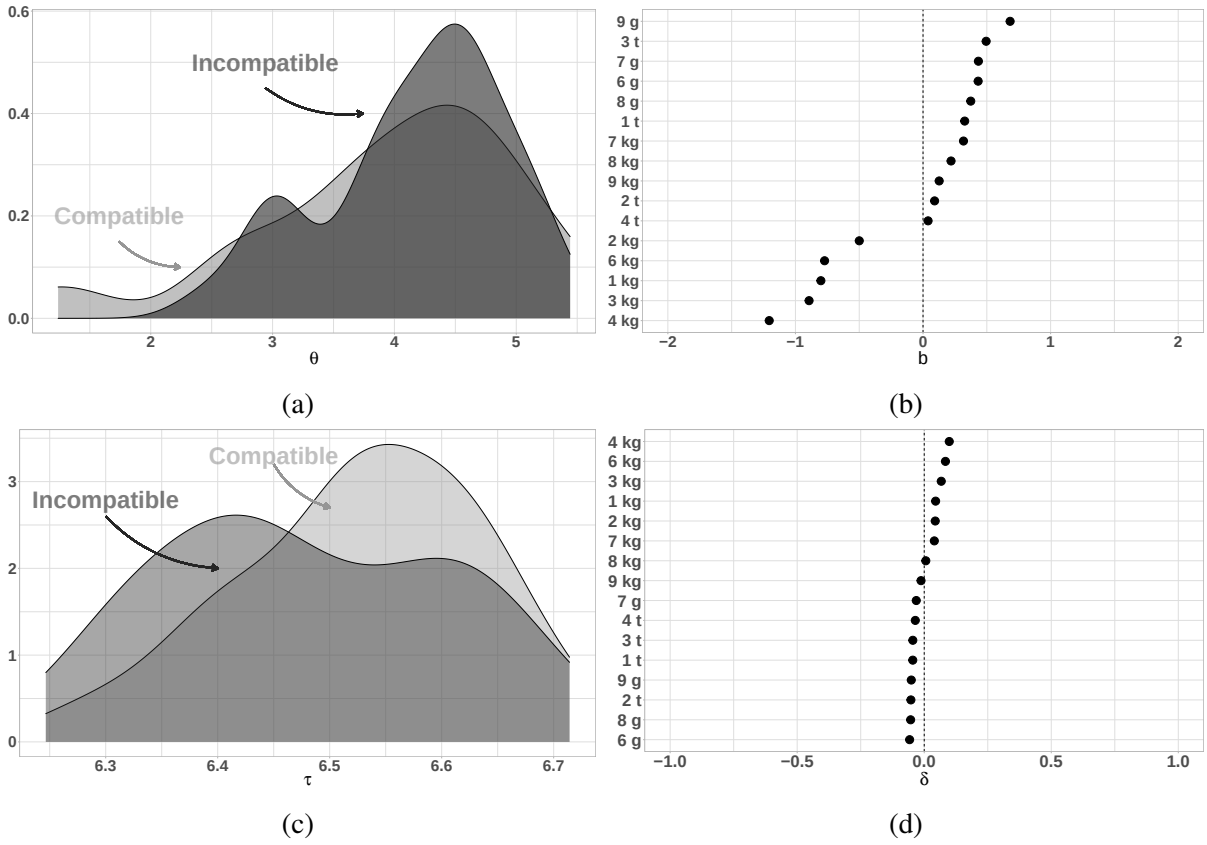


Figure 6: Panels 6a and 6b represent the distribution of the respondents' ability estimates in the two conditions and the stimulus easiness estimates across conditions of the Rasch-like model, respectively. Panels 6c and 6d represent the distribution of the respondents' speed and stimulus time intensity estimates of the log-normal model across the two contrasting conditions. The estimates are obtained from the SNARC-like data.

in Figure 6a:

```
# reshape the data frame with the speed estimates from wide to long
speed_long <- stack(speed_m3)
# rename the columns
colnames(speed_long) = c("tau", "condition")
# plot the density distribution of the speed estimates in the two conditions
ggplot(speed_long,
       aes(x = tau, fill = condition)) +
       geom_density()
```

The speed estimates are reported on the x-axis and the densities of τ in the two conditions are represented by the two curves of Figure 6c. The lower the value of τ, the higher the speed (i.e., the faster the observed responses). The speed estimates showed slightly higher variability in

one condition than in the other, and indicated that respondents tended to be faster in the former condition than in the latter one.

The graphical representation of the time intensity estimates δ_s (Figure 6d) are obtained with the same code used for representing the easiness estimates in Figure 6b with the appropriate replacements:

```
ggplot(intsensity_m3, aes(x = intensity, y = stimuli)) + geom_point()
```

The time intensity estimates δ_s are represented on the x-axis, while their corresponding labels are on the y-axis and a point is drawn in correspondence of each stimulus–estimate pair (Figure 6d). The lower the value of δ, the lower the amount of time required by the stimulus to get a response.

Implicit Association Test

The data pertains to an IAT experiment for the assessment of racial attitudes towards White and Black people (i.e., Race IAT), and it contains the observations on 65 respondents presented with 28 stimuli (12 object stimuli and 16 attribute stimuli). The stimuli representing four super-ordinate categories (two object categories, White people and Black people and two attribute categories, Good and Bad) are presented one at a time at the center of the screen and the respondents have to assign them to the correct category with two response keys in two contrasting conditions. The labels that identify the categories are displayed on the top-left and top-right sides of the screens, and the stimuli belonging to the categories on the same side of the screen are assigned with the same response key. In one condition (White-Good/Black-Bad condition, WGBB), White people faces and positive words are assigned with the same response key, while Black people faces and negative words are assigned with the opposite key. In the contrasting condition (Black-Good/White-Bad condition, BGWB), the labels of the White people and Black people categories switch positions on the screen, such that Black people faces and positive words are assigned with the same response key, and White people faces and negative words are assigned with the opposite key. The data, which are available at `https://ottaviae.github.io/DscoreApp/`, are published in Epifania et al. (2021) and they are the example data set of an online user friendly application for the computation of

the IAT *D* score (Epifania et al., 2020b).

The first rows of the IAT data set (object `mydata`) are displayed in Figure 7. The labels of the stimuli are reported in column `stimuli`, while the IAT contrasting conditions are reported in column `condition`. Response times and response accuracies to each stimulus in each condition are reported in columns `correct` and `latency`, respectively. The unique IDs of the respondents are stored in the column `participant`.

```
> head(mydata)
  participant condition  stimuli latency correct
1           1      WGBB     hate    1224       1
2           1      WGBB     bf14    5160       1
3           1      WGBB laughter    1214       1
4           1      WGBB     bf56    1143       1
5           1      WGBB     evil     827       1
6           1      WGBB      wf3    1859       1
```

Figure 7: Head of the Race IAT data set.

Accuracy and log-time models of the previous section are applied to the IAT accuracies and response times, respectively. Model comparison of the accuracy models is depicted in the top panel of Figure 8, while that of the log-time models is depicted in the bottom panel of Figure 8. Comparative fit indexes (AIC, BIC, log-likelihood, deviance) clearly indicated Model 3 as the best fitting model among the log-time models, while they did not straightforwardly suggest an accuracy model over another. AIC, deviance, and log-likelihood favored Model 2, while BIC suggested Model 1 as the best fitting one. Model 2 was chosen as the best fitting model for the response accuracies. As such, condition–specific speed and overall time intensity estimates of the log-normal model are obtained from the log-time model and condition–specific easiness and overall ability Rasch-like estimates are obtained from the accuracy model. The summary of the best fitting models can be obtained with the code illustrated in the previous section. Given that the same log-time model described in the previous section was selected as the best fitting model, it is no further illustrated nor commented in this section. The summary of the results of the model (`summary(iat_accuracy2)`) is illustrated in Figure 9.

The fixed effects pertaining to the two contrasting conditions were significantly different from 0. As the marginal log-odds suggest, one condition (3.4181) is expected to be easier than the other (2.0128) Concerning the random effects, respondents presented a good variability

```
> anova(iat_accuracy1, iat_accuracy2, iat_accuracy3)
Data: iat_data
Models:
iat_accuracy1: correct ~ 0 + condition + (1 | participant) + (1 | stimuli)
iat_accuracy2: correct ~ 0 + condition + (1 | participant) + (0 + condition | stimuli)
iat_accuracy3: correct ~ 0 + condition + (0 + condition | participant) + (1 | stimuli)
              npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
iat_accuracy1    4 4144.3 4172.1 -2068.1   4136.3
iat_accuracy2    6 4141.3 4183.1 -2064.7   4129.3 6.9271  2    0.03132 *
iat_accuracy3    6 4145.2 4187.0 -2066.6   4133.2 0.0000  0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(iat_time1, iat_time2, iat_time3)
Data: iat_data
Models:
iat_time1: log(latency) ~ 0 + condition + (1 | participant) + (1 | stimuli)
iat_time2: log(latency) ~ 0 + condition + (1 | participant) + (0 + condition | stimuli)
iat_time3: log(latency) ~ 0 + condition + (0 + condition | participant) + (1 | stimuli)
           npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
iat_time1    5 6073.2 6108.0 -3031.6   6063.2
iat_time2    7 6061.4 6110.2 -3023.7   6047.4  15.743  2  0.0003815 ***
iat_time3    7 5657.2 5705.9 -2821.6   5643.2 404.249  0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 8: Model comparison of the accuracy models (top panel) and of the log-time models (bottom).

```
> summary(iat_accuracy2)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: correct ~ 0 + condition + (1 | participant) + (0 + condition |      stimuli)
   Data: iat_data

     AIC      BIC   logLik deviance df.resid
  4141.3   4183.1  -2064.7   4129.3     7794

Scaled residuals:
    Min      1Q  Median      3Q     Max
-8.8764  0.1679  0.2242  0.3401  0.9156

Random effects:
 Groups      Name          Variance Std.Dev. Corr
 participant (Intercept)   0.26198  0.5118
 stimuli     conditionBGWB 0.13827  0.3718
             conditionWGBB 0.06751  0.2598   0.17
Number of obs: 7800, groups:  participant, 65; stimuli, 28

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
conditionBGWB   2.0128     0.1080   18.63   <2e-16 ***
conditionWGBB   3.4181     0.1226   27.88   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            cnBGWB
conditnWGBB 0.379
```

Figure 9: Summary of the results of Model 2 applied on the response accuracies of the IAT.

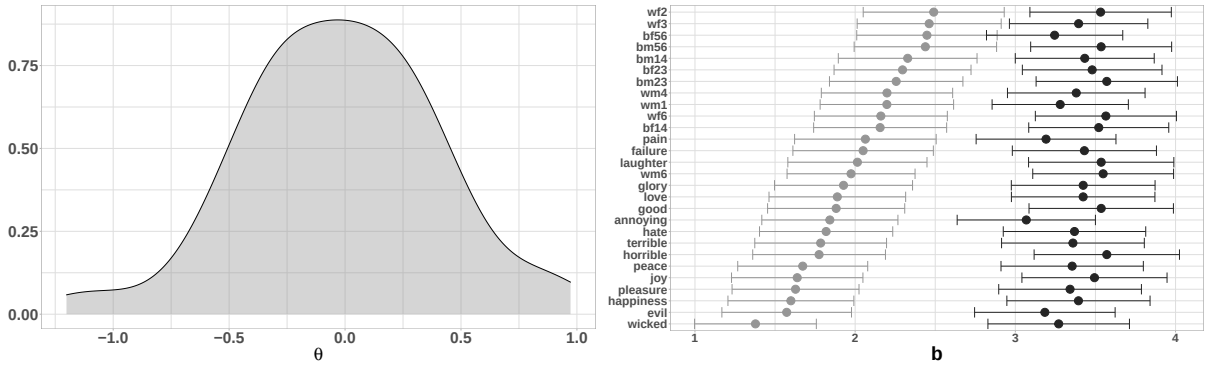
(0.26198), while the variability of the stimuli was higher in the one condition (0.13827) than in the opposite one (0.06751). Moreover, the correlation between the slopes of the stimuli in the

two conditions was small (0.17), this suggesting that the functioning of the stimuli varied a lot between conditions.

Figure 10a illustrates the distribution of the ability estimates obtained with the code:

```
ggplot(estimates_p_m1, aes(x = theta)) + geom_density()
```

where `estimates_p_m1` is a data frame containing the θ estimates of the respondents (represented on the x-axis). The condition–specific easiness estimates in Figure 10b are presented



(a) (b)

Figure 10: Panels 10a and 10b represent the distribution of the respondents' ability across the two conditions and the condition–specific stimulus easiness estimates of the Rasch-like model, respectively. The stimulus estimates in the BGWB condition are in light gray while those in the WGBB condition are in black.

with the 95% confidence intervals ($z = \pm 1.96$). The confidence intervals, which require the calculation of the conditional variance of the BLUP of each stimulus (Doran et al., 2007), can be obtained as:

```
# extract the BLUP of each stimulus along with the conditional variance
blup_easiness_iat = as.data.frame(ranef(m_accuracy1, condVar = TRUE))
# extract the fixed effects of the model and store them in a data frame
fixef_accuracy_iat = data.frame(term = names(fixef(m_accuracy1)),
                                est = (fixef(m_accuracy1)))
# merge together the BLUP and conditional variances of the stimuli and the fixed effects
easiness_iat = merge(blup_easiness_iat, fixef_accuracy_iat)
# rename the columun names
colnames(easiness_iat) = c("Condition", "grpvar", "stimuli", "blup", "se", "fix.est")
# compute the easiness estimates for each stimulus as the sum of the fixed effect and
```

```
# the BLUP of each stimulus
easiness_iat$b = with(easiness_iat, fix.est + blup)
```

The graphical representation in Figure 10b is obtained with:

```
ggplot(easiness_iat,
       aes(x = b, y = stimuli, color = Condition)) + geom_point() +
        geom_errorbarh(aes(xmin = b - 1.96 * se, xmax= b + 1.96 * se))
```

where the easiness estimates are on the x-axis (argument `x = b`), the labels of the stimuli are on the y-axis (argument `y = stimuli`), different colors are used for representing the conditions (argument `color = Condition`), and a circle is drawn in correspondence of each estimate–stimulus pair in each condition. The error bars are obtained with `geom_errorbarh()` by defining minimum (`xmin`) and maximum (`xmax`) values around the estimate of each stimulus.

As it can be seen from the graph, stimuli tended to be easier in WGBB condition than in the BGWB condition. Given that the intervals of the stimulus estimates in the two conditions are not overlapping but in one instance (i.e., stimulus bf56), the differences between the condition–specific easiness estimates are likely to be significant.

Final remarks

This manuscript presented a tutorial for obtaining a Rasch-like parametrization of accuracies and response times of data with fully-crossed design by exploiting the flexibility of LMMs and the variability in the data at different levels. An interesting take of the approach entails the detailed information that can be gathered at the single stimulus level by exploiting the sampling variability of the stimuli. For instance, Epifania et al. (2022a) used this approach for identifying the stimuli that mostly contributed to the IAT effect, as well as the stimuli that least contributed to it. Results in their study suggested that, by isolating the most contributing stimuli and computing the *D* score (i.e., the usual scoring for the IAT), it is possible to obtain a measure better able to predict a behavioral choice than the *D* score computed on all the stimuli.

Although the applications in this tutorial dealt with two experiments typical of cognitive psychology and implicit social cognition, the approach can be applied to any situation where the

respondents are presented multiple times with the same set of stimuli in at least two contrasting conditions.

Other approaches could have been used to analyze the response times, that do not require the log-transformation. For instance, they could have been analyzed considering a Gamma distribution with log or inverse link functions. Although these approaches are useful, they do not allow for obtaining respondent and stimulus estimates that are coherent with the log-normal model (van der Linden, 2006), which are in turn interpretable as Rasch-like estimates of the response times.

Since the focus of this manuscript was not on the formalization of the proposed approach, the model specification has only been briefly presented. Further details on the general approach can be found in De Boeck et al. (2011); Doran et al. (2007), while specific applications in implicit social cognition can be found in Epifania et al. (2022a); Epifania, Robusto, and Anselmi (2023); Epifania et al. (2021).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723. doi: https://doi.org/10.1109/TAC.1974.1100705

Anselmi, P., Vianello, M., & Robusto, E. (2011). Positive associations primacy in the IAT: A Many-Facet Rasch Measurement analysis. *Experimental Psychology*, *58*(5), 376–384. doi: https://doi.org/10.1027/1618-3169/a000106

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The Sorting Paired Features Task: A measure of association strengths. *Experimental Psychology*, *56*(5), 329–343. doi: 10 .1027/1618-3169.56.5.329

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *44*(3), 255–278. doi: https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: https://doi.org/10.18637/jss.v067.i01

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The Estimation of Item Response Models with the lmer Function from the lme4 Package in R. *Journal of Statistical Software*, *39*(12), 1–28.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371. doi: https://doi.org/10.1037/0096-3445.122.3.371

Doran, H., Bates, D., Bliese, P., & Dowling, M. (2007). Estimating the Multilevel Rasch Model : With the lme4 package. *Journal of Statistical Software*, *20*(2), 1–18. doi: https://doi.org/10.1111/j.1467-9868.2007.00600.x

Epifania, O. M., Anselmi, P., & Robusto, E. (2020a). A fairer comparison between the Implicit Association Test and the Single Category Implicit Association Test. *Testing, Psychomet-*

rics, Methodology in Applied Psychology, *27*(2), 207-220. doi: https://doi.org/10.4473/TPM27.2.4

Epifania, O. M., Anselmi, P., & Robusto, E. (2020b). DscoreApp: A Shiny Web Application for the Computation of the Implicit Association Test D-Score. *Frontiers in Psychology*, *10*, 2938. doi: https://doi.org/10.3389/fpsyg.2019.02938

Epifania, O. M., Anselmi, P., & Robusto, E. (2022a). Filling the gap between implicit associations and behavior: A Linear Mixed-Effects Rasch Analysis of the Implicit Association Test. *Methodology*, *18*(3), 185–202. doi: https://doi.org/10.5964/meth.7155

Epifania, O. M., Anselmi, P., & Robusto, E. (2022b). Implicit social cognition through the years: The Implicit Association Test at age 21. *Psychology of Consciousness: Theory, Research, and Practice*, *9*(3), 201-217. doi: https://doi.org/10.1037/cns0000305

Epifania, O. M., Robusto, E., & Anselmi, P. (2021). Rasch gone mixed: A mixed model approach to the Implicit Association Test. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, *28*(4), 467-483. doi: https://doi.org/10.4473/TPM28.4.5

Epifania, O. M., Robusto, E., & Anselmi, P. (2023). Is the performance at the Implicit Association Test sensitive to feedback presentation? A Rasch-based analysis. *Psychological Research*, *87*(3), 737-750. doi: https://doi.org/10.1007/s00426-022-01703-w

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*(6), 359–374.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Soclal Psychology*, *74*(6), 1464–1480. doi: https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. doi: https://doi.org/10.1037/0022-3514.85.2.197

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54-69. doi: https://

doi.org/10.1037/a0028347

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601–625. doi: https://doi.org/10.1146/annurev-psych-122414-033702

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

McCullagh, P., & Nelder, J. (1989). *Generalized linear models*. (2nd ed.). New York: Chapman & Hall.

Pastore, M. (2015). *Analisi dei dati in psicologia [data analysis in psychology]*. Il Mulino.

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. New York; New York: Springer Science & Business Media.

Rasch, G. (1960). *Probabilistic models for somo intelligence and attainment test*. Chicago, IL: The University of Chicago Press.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 461–464. doi: https://doi.org/10.1214/aos/1176344136

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181–204. doi: https://doi.org/10.3102/10769986031002181

Vicovaro, M., & Dalmaso, M. (2021). Is 'heavy'up or down? testing the vertical spatial representation of weight. *Psychological Research*, *85*(3), 1183–1200. doi: https://doi.org/10.1007/s00426-020-01309-0

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5). doi: https://doi.org/10.1037/xge0000014

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling stimulus variation in three common implicit attitude tasks. *Behavior Research Methods*, *49*(4), 1193–1209. doi: https://

doi.org/10.3758/s13428-016-0779-0

Zhang, P., Cao, B., & Li, F. (2022). The role of cognitive control in the SNARC effect: A review. *PsyCh Journal*. doi: https://doi.org/10.1002/pchj.586