



Testing Interactions in Multinomial Processing Tree Models

Beatrice G. Kuhlmann¹, Edgar Erdfelder^{1*} and Morten Moshagen²

¹ Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany, ² Department of Psychology, Ulm University, Ulm, Germany

Multinomial processing tree (MPT) models allow testing hypotheses on latent psychological processes that underlie human behavior. However, past applications of this model class have mainly been restricted to the analysis of main effects. In this paper, we adopt the interaction concept as defined in log-linear models and show why it is appropriate for MPT models. We then explain how to implement and test ordinal and disordinal two-way interaction hypotheses in MPT models. We also show how our method generalizes to higher-order interactions involving three or more factors. An empirical example from source memory and aging demonstrates the applicability of this method and allows for directly testing the associative deficit theory that age differences are larger in associative (e.g., source) memory as opposed to item memory. Throughout the paper, we explain how most analytic steps can be easily implemented in the freely available software multiTree.

OPEN ACCESS

Edited by:

Hans Colonius,
University of Oldenburg, Germany

Reviewed by:

Oscar Lorenzo Olvera Astivia,
University of South Florida,
United States

Rubén Maneiro,
Pontifical University of
Salamanca, Spain

*Correspondence:

Edgar Erdfelder
erdfelder@
psychologie.uni-mannheim.de

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 31 July 2019

Accepted: 03 October 2019

Published: 01 November 2019

Citation:

Kuhlmann BG, Erdfelder E and
Moshagen M (2019) Testing
Interactions in Multinomial Processing
Tree Models. *Front. Psychol.* 10:2364.
doi: 10.3389/fpsyg.2019.02364

Keywords: multinomial processing tree models, interactions, parametric order constraints, associative deficit hypothesis, cognitive aging

INTRODUCTION

Psychologists are typically interested in internal processes (e.g., cognitions and emotions) that drive behavior but are not directly observable. Multinomial processing tree (MPT) models are stochastic models that, based on observable participant responses, allow for estimation of the probabilities of such unobservable processes taking place or not. Developed in the 1980s and 1990s (Batchelder and Riefer, 1980, 1986, 1999; Riefer and Batchelder, 1988; Hu and Batchelder, 1994), they are currently widely used in several branches of psychological research (see Erdfelder et al., 2009, for a comprehensive review). However, so far almost all applications of MPT models in psychology involve simple parameter comparisons across experimental conditions or groups of participants, thus testing only main effects on model parameters, arguably because standard MPT parameter tests do not readily allow to test interactions. In the present paper, we explain how MPT models can be reparameterized to allow for testing interaction hypotheses. We provide an easy-to-follow introduction and an application example from cognitive aging on how to implement parameter constraints to test two-way (and higher-order) interaction hypotheses, using the software multiTree (Moshagen, 2010).

A Brief Introduction to MPT Models

MPT models are used to analyze categorical data such as response frequencies. In contrast to general-purpose data analysis techniques for categorical data (e.g., log-linear models, Read and Cressie, 1988; Agresti, 2002), MPT models are tailored to a specific psychological research paradigm and are based on substantive theoretical work regarding the processes involved in the psychological

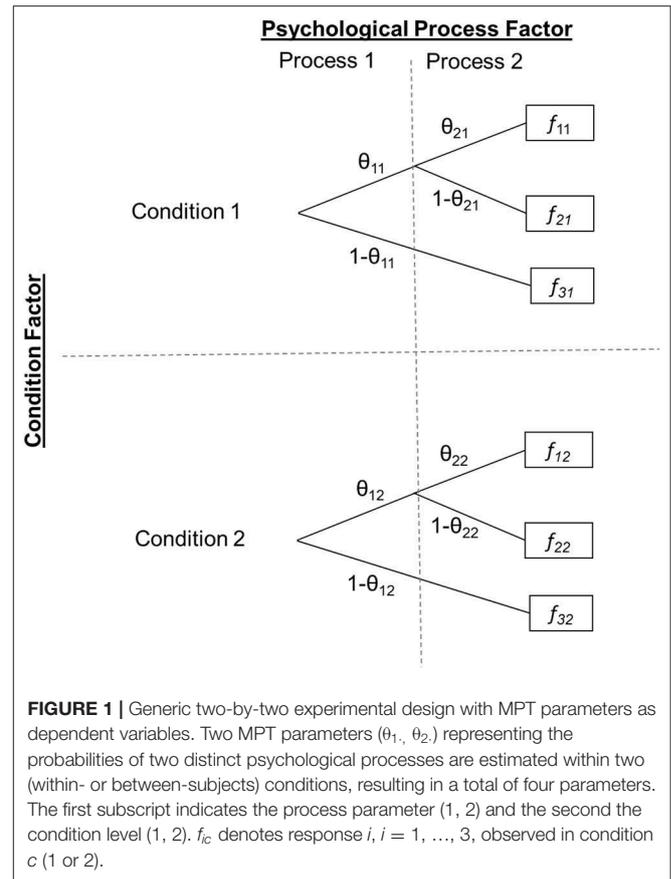
phenomenon investigated in this paradigm (e.g., Bayen et al., 1996; Rummel et al., 2011). Thereby, MPT models can be used to evaluate theories of a given psychological phenomenon as well as to estimate probabilities of the latent processes specified in the theory. In this section, we provide a brief introduction to MPT models and standard inferential tests available for MPT model parameters (for technical details, see Riefer and Batchelder, 1988; Hu and Batchelder, 1994; Batchelder and Riefer, 1999).

The upper half of **Figure 1** displays a very simple generic MPT model containing two parameters, θ_1 and θ_2 . For now we will ignore the second subscript, which indicates that the model and its two parameters are estimated separately for two groups (e.g., experimental conditions). The boxes to the right indicate observable categorical participant responses in a psychological paradigm, for example “Remember,” “Know,” and “New” responses in a Remember-Know recognition task (Gardiner, 1988), with three distinct responses in total (f_1 , f_2 , and f_3 again observed separately per condition). The core idea of any MPT model is to relate the probabilities of such observable responses to the probabilities of unobservable psychological processes (measured by parameters θ_1 and θ_2). Specifically, in this simplified generic model¹, the three observable response probabilities relate to the model parameters as follows:

$$\begin{aligned} p(f_1) &= \theta_1 \cdot \theta_2 \\ p(f_2) &= \theta_1 \cdot (1 - \theta_2) \\ p(f_3) &= (1 - \theta_1) \end{aligned} \quad (1)$$

To estimate these parameters, MPT models assume that the frequencies of the observable responses follow a (product) multinomial distribution. This distribution is then described in terms of the model parameters by replacing the multinomial response probabilities with the right side of the corresponding model equation. Thus, any parameter in the model is an unknown probability varying in the real interval $[0, 1]$. Parameter estimates are typically obtained by means of maximum likelihood estimation, more specifically, by minimizing the log-likelihood ratio goodness-of-fit statistic $G^2(df)$ using the expectation-maximization algorithm (Hu and Batchelder, 1994). Given that the model holds and certain regularity conditions are met (Read and Cressie, 1988, Chapter 4), the resulting minimum $G^2(df)$ is asymptotically $\chi^2(df)$ distributed, with the degrees of freedom (df) corresponding to the number of independent response probabilities minus the number of estimated parameters. The obtained model fit statistic can thus be compared to a $\chi^2(df)$ distribution to evaluate the null hypothesis (H_0) that the model fits the data. The H_0 of model fit is rejected when the observed $G^2(df)$ fit statistic falls in the upper $\alpha \cdot 100\%$ of the corresponding $\chi^2(df)$ reference distribution (typically, $\alpha = 0.05$ or $\alpha = 0.01$) and retained otherwise. Several computer programs are freely available for model fitting and parameter estimation (Hu and Phillips, 1999; Stahl and Klauer, 2007; Moshagen, 2010; Singmann and Kellen, 2013; Heck et al., 2018).

¹For a more elaborated MPT model of the processes involved in the Remember-Know recognition task see Erdfelder et al. (2007).



To test more specific hypotheses concerning the model parameters, restrictions can be imposed reflecting the H_0 that a model parameter is either equal to (1) a constant value (e.g., chance level for guessing between two options = 0.50) or (2) another parameter in the model. It is then tested whether this restriction significantly worsens model fit by comparing the fit of the restricted model [$G^2_r(df_r)$] to the fit of the unrestricted model [$G^2_u(df_u)$], based on a χ^2 difference test [i.e., $\Delta G^2(df_r - df_u) = G^2_r(df_r) - G^2_u(df_u)$]. If model fit is significantly worsened by the restriction, the H_0 that the parameter restriction holds is rejected.

However, neither of these standard MPT parameter tests allows for testing interactions in multifactorial designs. To illustrate this problem, consider the two-factorial experimental design exemplified in **Figure 1**. First, there is a Condition factor with two levels (Condition 1, Condition 2). Our generic MPT model is assumed to hold in each of these two conditions with presumably different parameter values. Second, there is a Process factor representing the psychological mechanism of interest, again with two levels (θ_1 , θ_2). A standard MPT parameter test can be used to test for a main effect of the Condition factor on the psychological processes by restricting θ_1 and θ_2 to be equal across conditions ($\theta_{11} = \theta_{12}$ and $\theta_{21} = \theta_{22}$). Alternatively, one can test for a main effect of the psychological process by simultaneously restricting θ_1 and θ_2 to be equal within each condition ($\theta_{11} = \theta_{21}$ and $\theta_{12} = \theta_{22}$). Additionally, the simple main effects of the

Process factor can be assessed for each condition separately or, if more appropriate, the simple main effects of the Condition factor separately for each psychological process. However, all these tests do not straightforwardly implement the test of an interaction of the Condition factor with psychological process. For example, a theory might predict that there should be a general condition main effect with Condition 1 outperforming Condition 2 but that this effect is moderated by the type of psychological process, such that the condition effect is more strongly pronounced in the process represented by parameter θ_2 than in the process represented by θ_1 . In this case, one could test the effect of condition on both parameters but would lack a test for comparing the strengths of this condition effect between the two processes.

METHOD

In the following, we will derive a method to test interactions on MPT model parameters. We will primarily focus on two-way interactions but later explain how our method generalizes to testing higher-order interactions. For now, we will assume the experimental design with two factors illustrated in **Figure 1** comprising a Condition factor with two levels (Condition 1, Condition 2; within- or between-subjects) and a Process factor with two levels (Parameter θ_1 , Parameter θ_2). Note that, in principle, the second factor could also reflect any different grouping (such as an experimentally manipulated within or between condition factor). In this case, only one MPT model parameter would be of interest that is estimated in the four cells resulting from fully crossing both condition factors.

Appropriate Interaction Concept for MPT Models

When assessing interactions, a distinction needs to be made between interactions as defined in analysis of variance (ANOVA) models and interactions as defined in Log-Linear Models (LLM). In our Condition \times Process design, a null interaction in the ANOVA sense would refer to invariance of parameter *differences* between the levels of one factor across the levels of the second factor. That is, a null interaction would mean that the difference, δ , between conditions is identical on both parameters. Likewise, the difference between the two process parameters would need to be the same across conditions. In the following, we will demonstrate that due to the nature of MPT model parameters, which represent probabilities and are thus bounded by 0 and 1, this linear interaction concept is not suited for this type of models and may falsely imply the presence of an interaction in the presence of strong main effects. To demonstrate this, let us assume, without loss of generality, that factor levels are arranged such that moving from Level 1 to Level 2 on either factor results in a constant decrease of the corresponding parameters under the null interaction model. Therefore, the following would need to hold for a null interaction of condition and process in the ANOVA sense:

$$\theta_{11} - \theta_{12} = \theta_{21} - \theta_{22} = \delta_C \quad (2)$$

$$\theta_{11} - \theta_{21} = \theta_{12} - \theta_{22} = \delta_P \quad (3)$$

with δ_C and δ_P indicating the main effects of the two factors [Condition (C), Process (P)], $0 \leq \delta_C \leq 1$ and $0 \leq \delta_P \leq 1$. Therefore, under the null hypothesis of no interaction, parameter θ_{22} would be affected by both main effects as follows

$$\theta_{22} = \theta_{11} - \delta_C - \delta_P. \quad (4)$$

Because θ_{22} , like all MPT parameters, represents a probability, it must lie within $[0, 1]$. Consequently, the following restriction would be imposed on the main effects of condition and process for a null interaction to be present:

$$(\delta_C + \delta_P) \leq \theta_{11} \quad (5)$$

A weird consequence of (5) is that strong main effects would automatically imply a violation of (5) and thus an interaction in the ANOVA sense. If, for example, $\theta_{11} = 0.70$ and $\theta_{12} = 0.10$ (i.e., $\delta_C = 0.60$), an interaction of condition and process would automatically be implied whenever the Process factor has a main effect larger than $\delta_P = 0.10$ to ensure that $\theta_{22} \geq 0$. More generally, if the largest parameter is quite small (e.g., 0.05) not much shrinkage is possible if defined as a difference (i.e., $0 \leq \delta \leq 0.05$). In the standard ANOVA framework, in contrast, main effects and interactions could vary independently because parameters are not bounded by 0 and 1.

This obvious problem can be avoided in MPT modeling by adopting the LLM interaction concept in which a null interaction refers to invariance of parameter *ratios* across the levels of the second factor. That is, in terms of the 2×2 model with a Condition and a Process factor, a null interaction in the LLM sense would imply:

$$\theta_{12}/\theta_{11} = \theta_{22}/\theta_{21} = \alpha_C \quad (6)$$

$$\theta_{21}/\theta_{11} = \theta_{22}/\theta_{12} = \alpha_P \quad (7)$$

In this case, assuming no LLM-type interaction, θ_{22} would be:

$$\theta_{22} = \theta_{11} \cdot \alpha_C \cdot \alpha_P \quad (8)$$

Crucially note that, in order to keep all MPT parameter values within $[0, 1]$, main effects must be represented as ratios α_C and α_P with the larger parameter values always appearing in the denominator. Most importantly, however, an interaction effect is not automatically implied when α_C and α_P take on extreme values in $[0, 1]$, because main effects are multiplicative rather than additive under the null interaction hypothesis in the LLM sense. Thereby, even strong main effects can occur (i.e., α parameters close to 0) without implying an interaction. For the example given earlier, if $\theta_{11} = 0.70$ and $\theta_{12} = 0.10$ (i.e., $\alpha_C = 0.14$), the difference $\theta_{21} - \theta_{11}$ could still exceed 0.10 without necessarily implying an interaction. For example, if $\theta_{12} = 0.40$ (i.e., $\alpha_P = 0.57$), a null interaction in the LLM sense would require $\theta_{22} = 0.06$, a valid parameter value. Thus, main effects and interactions would not be artificially confounded. Likewise, the α parameters are not restricted by the value of the original parameter; even if the largest parameter is quite small, the α parameter can, in principle, be of any value between 0 and 1. Hence, we will adopt the LLM

interaction concept, because it is better tailored to multinomial models. For brevity, we refer to this latter concept of “interaction as defined in LLM” simply as “interaction” in what follows².

Estimating MPT Parameter Ratios via Parametric Order Constraints

Testing an interaction hypothesis essentially entails comparing a factor’s simple main effect under different conditions. As reasoned above, it is crucial that the main effects are expressed as *parameter ratios* (rather than differences) to test an interaction in the LLM sense. In order to perform this type of comparison with the parameter tests available in MPT models, it is thus necessary to introduce parameters that quantify a factor’s simple main effect in each possible condition as a ratio. Then, these ratio parameters can be compared across conditions to assess whether an interaction is present (i.e., the factor’s simple main effect differs depending on conditions) or not.

Knapp and Batchelder (2004) introduced reparameterizations of MPT models that allow estimating parameter ratios via *parametric order constraints*. As described earlier, all estimated MPT model parameters vary in [0, 1] when freely estimated but can be restricted to be equal to a constant or to another parameter. In contrast, parametric order constraints allow to implement restrictions that still let the constrained parameter vary freely but only within a restricted range of values. For example, the restriction $0 \leq \theta_1 \leq \theta_2 \leq 1$ constrains the parameter θ_1 to vary only in $[0, \theta_2]$ instead of $[0, 1]$, with the upper boundary being determined by the value of another model parameter, θ_2 . Once a parametric order constraint, such as $\theta_1 \leq \theta_2$, has been defined and a MPT model has been reparameterized to reflect this constraint, a new parameter α is estimated that reflects the shrinkage of θ_1 relative to θ_2 . Crucially, we propose that this shrinkage parameter α can be used to test interactions because it reflects the *relative* change in parameters and thus, in essence, represents the ratio $\alpha = \theta_1/\theta_2$.

Knapp and Batchelder (2004) introduced two equivalent reparameterization methods for MPT models to assess parametric order constraints. We will focus on their Method A, which has been implemented in the software multiTree (Moshagen, 2010). This method replaces a parameter θ by a novel parameter α such that α reflects *decreases* in θ , allowing the implementation of a constraint that restricts one parameter (e.g., θ_1) to be *smaller than or equal to* a second parameter (e.g., θ_2). Note that such a reparameterization does not change the dimensionality of the model (i.e., the number of parameters) and thus does not change the *df* of the model test. Also, note that this reparameterization is data-equivalent to the original model whenever the implemented order constraint actually holds in the observed data.

²Because invariant parameter ratios correspond to invariant differences between the logarithms of the parameters, a technically more accurate term would be “interaction on a logarithmic scale” (cf. Klauer et al., 2000, p. 861). In fact, Klauer and collaborators already employed the LLM concept of no interaction in the context of their belief bias MPT model (Klauer et al., 2000), albeit without discussing its implementation in MPT models and its implications for MPT modeling in general.

To apply this method for purposes of testing interactions, data from at least two (within- or between-subjects) conditions are required across which a parameter may change. Returning to our two-factorial example illustrated in **Figure 1**, there is a Condition factor with two levels/groups (e.g., younger and older adults). As before, let us assume that Condition 1 outperforms Condition 2 on both psychological processes measured by the model’s parameters θ_{11} and θ_{21} . Via Knapp and Batchelder’s method A, we can replace θ_{12} and θ_{22} by two new parameters, α_{C,θ_1} and α_{C,θ_2} , representing the shrinkage factors for each process in Condition 2 compared to Condition 1. In the reparameterized model, the parameters of Condition 2 are thus reparameterized as follows:

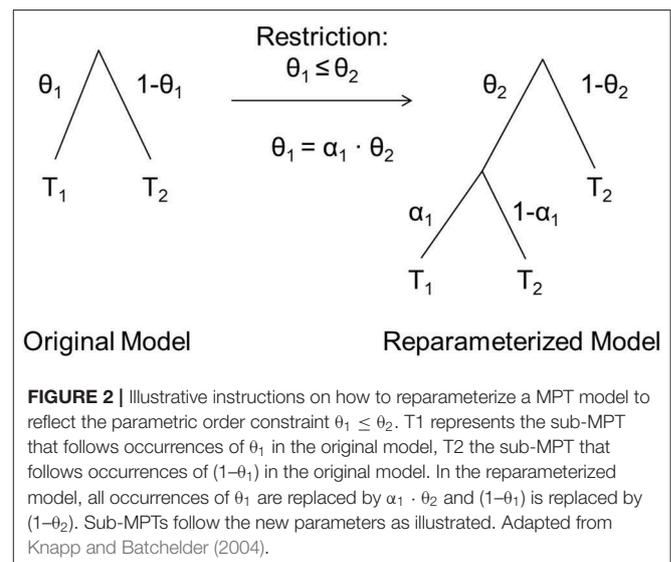
$$\theta_{12} = \theta_{11} \cdot \alpha_{C|\theta_1} \tag{9}$$

$$\theta_{22} = \theta_{21} \cdot \alpha_{C|\theta_2} \tag{10}$$

Importantly, each occurrence of θ_{12} and θ_{22} in the original model is replaced with the corresponding product. The introduction of two new parameters α_{C,θ_1} and α_{C,θ_2} to the model of course also requires the introduction of the corresponding complementary branches $(1-\alpha_{C|\theta_1})$ and $(1-\alpha_{C|\theta_2})$. **Figure 2** shows how the reparameterized model can be derived from the original model. Whereas the equations for Condition 1 correspond to those presented in Equation (1) except for an additional index “1”, for Condition 2 the reparameterized model equations become:

$$\begin{aligned} p(f_{12}) &= \theta_{11} \cdot \alpha_{C|\theta_1} \cdot \theta_{21} \cdot \alpha_{C|\theta_2} \\ p(f_{22}) &= \theta_{11} \cdot \alpha_{C|\theta_1} \cdot \theta_{21} \cdot (1 - \alpha_{C|\theta_2}) + \theta_{11} \cdot \alpha_{C|\theta_1} \cdot (1 - \theta_{21}) \\ p(f_{32}) &= \theta_{11} \cdot (1 - \alpha_{C|\theta_1}) + (1 - \theta_{11}) \end{aligned} \tag{11}$$

Obviously, the reparameterized model will always be larger (in terms of the number of branches) than the original model. Therefore, an automatized implementation of order constraints is recommended not only because it is less tedious but, more importantly, less error-prone. To our knowledge, multiTree is



the only MPT modeling software that offers this automatic reparameterization of MPT models to reflect parametric order constraints. Apart from this, the reparameterized model is equivalent to the original model for the subspace of the parameter space fulfilling the implemented order constraints (cf., Meiser, 2014). Thus, the properties (e.g., identifiability; validity) of the original model, already shown by Bayen et al. (1996), also hold for the reparameterized model. Crucially note that this only pertains if the reparameterization is implemented correctly, that is if the larger parameter is in the denominator of the ratio. Indeed, in this case, and further given that the ratio parameters are not at the boundary of parameter space (i.e., 0 or 1), model fit of the reparameterized model will be identical to the original model.

Note that even though this reparameterization restricts θ_{12} and θ_{22} to vary only in $[0, \theta_{11}]$ and $[0, \theta_{21}]$, respectively, all parameters in the reparameterized model may vary in $[0, 1]$, so that standard MPT procedures can be used to estimate and test the model. This is because the restricted parameters, θ_{12} and θ_{22} , are not included in the reparameterized model and the new parameters, $\alpha_{C|\theta_1}$ and $\alpha_{C|\theta_2}$, reflect shrinkage factors (i.e., $\alpha_{C|\theta_1} = \theta_{12}/\theta_{11}$ and $\alpha_{C|\theta_2} = \theta_{22}/\theta_{21}$) that can vary in $[0, 1]$. If $\alpha_{C|\theta_1} = \alpha_{C|\theta_2} = 1$, then there is no condition effect on both process parameters, that is $\theta_{12} = \theta_{11}$ and $\theta_{22} = \theta_{21}$. In contrast if $\alpha_{C|\theta_1} < 1$ and $\alpha_{C|\theta_2} < 1$, then Condition 2's parameters are smaller than Condition 1's for both psychological processes of interest.

In case of a design including factors with more than two levels (e.g., three or more different psychological process parameters of interest), one would specify additional ratio parameters. For example, with three levels of the Process factor ($\theta_1, \theta_2, \theta_3$) and two conditions, there would be three simple main effects of Condition 1 vs. 2 ($\alpha_{C|\theta_1}, \alpha_{C|\theta_2}, \alpha_{C|\theta_3}$). Similarly, when there are more than two conditions, one would specify additional shrinkage parameters. In general, if there are $K \geq 2$ levels of the Condition factor, then the number of shrinkage parameters required per process parameter would be $K-1$. For example, in a design with two process parameters and three conditions, one would have four shrinkage parameters in total, two for each of the two processes. The two shrinkage factors per process could, for example, represent the shrinkage in a process parameter from Conditions 1 to 2 and from Conditions 2 to 3 or, alternatively, the shrinkage from Conditions 1 to 2 and from Conditions 1 to 3. The first method would be appropriate if one expects a main effect decrease from Conditions 1 to 2 and, additionally, from Conditions 2 to 3. In contrast, if one simply expects Condition 1 to perform best with no specific ordering of Conditions 2 and 3, then one would select the second pattern of order constraints. Although the values of the second shrinkage parameter will in general differ between both methods, they are equivalent in the sense that they will yield identical interaction test results, provided that the imposed order constraints are in agreement with the observed data.

So far, we have assumed that the process effect is in the same direction across all conditions. In this case, an *ordinal* interaction would be present if this process effect significantly differs in size across conditions. In contrast, a *disordinal* interaction would be present if a factor has opposite effects across at least two levels of the other factor (also called “crossover interaction” or

“double dissociation”). For example, if the Condition factor in our example from **Figure 1** produced a decrease from the first to the second level in the first process parameter (i.e., $\theta_{12} < \theta_{11}$) but an increase in the second parameter (i.e., $\theta_{22} > \theta_{21}$) then this would establish a disordinal interaction. In turn, we will explain how to test both types of interactions.

Testing Ordinal Two-Way Interactions in MPT Model Parameters

In ordinal interactions, the effect of one factor is in the same direction for each level of the other factor while it may vary in effect size (including zero effects). In our example, Condition 2 may always have lower process probabilities than Condition 1 irrespective of the type of process (θ_1, θ_2) but this difference may be more pronounced for one process than the other. To test such an interaction, one would implement the shrinkage parameters as above to reflect the effect in the expected direction at each level of the other factor. When the observed data satisfy all imposed order constraints (i.e., when an increase in the level of the Condition factor is always associated with a decrease in all relevant process parameter estimates), the model fit for the original and the reparameterized model is necessarily identical. To formally test whether there is an ordinal interaction, one then restricts the corresponding shrinkage parameters to be equal (i.e., $\alpha_{C|\theta_1} = \alpha_{C|\theta_2}$). If this restriction significantly worsens model fit, that is, if the observed $\Delta G^2(df)$ falls in the upper $\alpha \cdot 100\%$ of the reference distribution, using the asymptotic $\chi^2(df)$ distribution as a reference with $df =$ number of equality constraints (i.e., $df = 1$ in our example), the H_0 assuming no interaction (i.e., identical shrinkage parameters across all levels of the other factor) would be rejected in favor of the H_1 that there is an ordinal two-way interaction.

Assuming that the interaction test turned out to be not significant, the H_0 of no interaction is maintained. The next step would be to test the main effect of the Condition factor (restricted to be equal across processes, i.e., $\alpha_{C|\theta_1} = \alpha_{C|\theta_2} = \alpha_C$) for significance. The most straightforward way to do this would be to test $H_0: \alpha_C = 1$ using a $\Delta G^2(df)$ difference test with respect to the null interaction model. However, since this H_0 lies at the boundary of the parameter space, the regularity conditions for the asymptotic $\chi^2(df)$ test are not met (Read and Cressie, 1988). This problem can be remedied using the parametric bootstrap option of multiTree (Moshagen, 2010) that generates the relevant reference distribution under H_0 empirically using Monte-Carlo methods (Efron and Tibshirani, 1993).

If, in contrast, the interaction effect turns out to be significant, it will often be of interest to test whether the interaction is weakly ordinal with respect to a certain factor (e.g., $\theta_{12} < \theta_{11}$ and $\theta_{22} = \theta_{21}$, also called “simple dissociation”) or strictly ordinal (e.g., $\theta_{12} < \theta_{11}$ and $\theta_{22} < \theta_{21}$, also called “no dissociation”), that is, whether each of the simple main effects is significant or not. This can be done in the same way as previously described for the test of the main effect null hypothesis $H_0: \alpha_C = 1$. This time, however, one would test the simple main effects separately, for example, $H_0: \alpha_{C|\theta_1} = 1$ or $H_0: \alpha_{C|\theta_2} = 1$. Again, we recommend the parametric bootstrap to evaluate statistical significance.

Note that if one or both factors have three or more levels and there are no specific predictions concerning which specific combination of the two factors should differ from the other design cells, one should first conduct an omnibus test by equating all shrinkage parameters. If this effect is significant, resulting in an acceptance of the overall interaction hypothesis, one can then follow up with pairwise comparisons of the shrinkage parameters to describe this interaction in more detail. We recommend adjusting such exploratory multiple pairwise comparisons using the Bonferroni-Holm method (Holm, 1979) to prevent alpha inflation. If there are more specific a priori predictions concerning which design cells should differ from the others, then only the corresponding shrinkage parameter constraints should be tested directly to minimize the overall number of tests.

Notably, presence vs. absence of an interaction between two factors of course does not depend on the order in which these factors enter into the interaction test. Applied to our example, rather than testing the null interaction hypothesis as outlined above, that is, by equating the shrinkage parameters representing simple condition main effects (i.e., $\alpha_{C|\theta 1} = \alpha_{C|\theta 2} = \alpha_C$), we could also invert our reparameterization scheme and equate shrinkage parameters representing simple process main effects within conditions (i.e., $\alpha_{P|C1} = \alpha_{P|C2} = \alpha_P$). Both constraints are equivalent and thus lead to the same null interaction model. By implication, the goodness-of-fit statistics $\Delta G^2(1)$ and test results will be identical for both types of constraint. To see the equivalence, note that $\alpha_{C|\theta 1} = \alpha_{C|\theta 2} = \alpha_C$ can be rewritten as:

$$\theta_{12} = \theta_{11} \cdot \alpha_{C|\theta 1} = \theta_{11} \cdot \alpha_C \quad (12)$$

$$\theta_{22} = \theta_{21} \cdot \alpha_{C|\theta 2} = \theta_{21} \cdot \alpha_C \quad (13)$$

Dividing (13) by (12) immediately leads to

$$\theta_{22}/\theta_{12} = \theta_{21}/\theta_{11} \iff \alpha_{P|C2} = \alpha_{P|C1} = \alpha_P. \quad (14)$$

Testing Disordinal Two-Way Interactions in MPT Model Parameters

In disordinal interactions, the direction of the effect of one factor differs across the levels of the other factor. If such an interaction holds, then a model imposing order constraints in the same direction across all levels of another factor will not fit the observed data perfectly. However, not all cases with order-constrained models resulting in $G^2 > 0$ automatically justify acceptance of the disordinal interaction hypothesis. One possibility is that the levels of the Condition factor were just all specified in the wrong order. By re-arranging the order of factor levels one could perhaps have an order-constrained model with perfect fit to the data. This option should be checked first by inspecting raw parameter estimates for the unconstrained model. If re-arranging factor levels results in perfect fit of the order-constrained model one should proceed with this revised order model and test for an ordinal interaction as described in the previous section. Of course, such a change in the order constraints relative to the originally expected order must be reported in any publication on these data.

If inspection of the data shows that misfit of the order-constrained model cannot be avoided by re-arranging the order of factor levels, the possibility remains that a weakly ordinal interaction underlies the data but sampling error caused a disordinal interaction pattern in the observed data. Given this, the H_0 of a weakly ordinal interaction (represented by the order-constrained model without any further constraints) cannot be evaluated using the asymptotic $\chi^2(df)$ as a reference distribution (Read and Cressie, 1988). However, one can again use the parametric bootstrap to estimate the limiting distribution of G^2 under the H_0 of an ordinal interaction. If the reparameterized model with shrinkage parameters in line with the ordinal interaction hypothesis fits the data, indicated by a bootstrapped p -value larger than the desired alpha-level, then the H_0 of an ordinal interaction is maintained.

Again, in the case of three or more levels per factor one should either only test pairwise comparisons that were a priori hypothesized to differ or conduct an omnibus restricting all ratio parameters to be equal first and proceed with pairwise tests of all possible combinations only if the omnibus test is significant. As before, we recommend applying a Bonferroni-Holm correction of the alpha significance level, especially if many levels are compared.

In contrast, if the bootstrapped p -value is equal to or smaller than the specified alpha-level, then the H_0 of an ordinal interaction is rejected. In other words, there are substantial effects in opposing directions, thus establishing a disordinal interaction. Again, this approach can be easily extended to factors with three or more levels by specifying as many ratio parameters as there are levels.

RESULTS

We now illustrate our method to test a prominent theory of cognitive aging that predicts an ordinal two-way interaction of age group and type of memory on memory performance. Specifically, the associative deficit hypothesis (Naveh-Benjamin, 2000) maintains that aging particularly impacts memory for associations such that differences between younger and older adults are most pronounced on tests of associative memory but less so on tests of simple item memory. One type of associative memory is source memory (memory for the context in which item information was first learned), such as determining whether we read information in a reliable or unreliable newspaper (Johnson et al., 1993). Thus, source memory involves memory for the association of an item to perceptual (e.g., font type) and spatio-temporal context features that together make up the source or origin of the item and should thus be more affected by aging than memory for the item itself.

In the typical source-monitoring paradigm, participants study items (e.g., words) presented in one of two different sources (e.g., in bold vs. italic text type). Subsequently, participants are tested with a list of previously studied (i.e., old) items intermixed with unstudied (i.e., new) items, all presented in a source-neutral manner (e.g., a regular text type). For each test item, participants are required to provide an old vs. new judgment. If an item is judged as old, they also have to indicate the source of the item.

A well-validated MPT model commonly used for this paradigm is the two-high threshold (2HT) model of source monitoring (Bayen et al., 1996; see Bröder and Meiser, 2007, for a review of alternative MPT models). A simple variant (Submodel 4) of this model is depicted as a tree structure in **Figure 3**. There are three trees, one for test items that were originally studied in Source A (e.g., italic text type), another for test items originally studied in Source B (e.g., bold text type), and an additional tree for new distractor items. The model reflects the theoretical assumption that memory and guessing processes operate together in a source-monitoring task (cf., Johnson et al., 1993). That is, a correct source attribution at test for items originally studied in one of the two sources may result from actual memory for both the item (probability D) and the source (probability d), or from a series of guessing processes in the absence of memory. That is, if participants remember an item (probability D) but not its source (probability $1-d$), they guess “Source A” with probability g and “Source B” with the complementary probability $1-g$. In a state of uncertainty about the item (probability $1-D$), the guessing process b captures the probability of guessing “old” and the complementary probability $1-b$ reflects the probability of guessing “new.” For new items, parameter D reflects the probability of distractor detection. In a state of uncertainty about the new item’s status (with probability $1-D$), item and, if applicable, source guesses follow as they do for previously studied but unrecognized items.

Thus, the 2HT-MPT model of source monitoring provides measures of item memory (parameter D) and source memory (parameter d) that are unconfounded by inferential guessing processes. These process parameters are thus ideal dependent variables to test the age group by memory type interaction postulated by the associative-deficit hypothesis.

Using younger and older adults’ source monitoring data from Kuhlmann and Touron (2011; all-once condition) we will now illustrate how to implement our method delineated above to test the ordinal interaction predicted by the associative-deficit hypothesis. In Kuhlmann and Touron’s study, 30 younger (mean age = 18.5 years) and 30 older (mean age = 67.2 years) adults studied 50 words, 25 appearing in italic text type and the remaining 25 appearing in bold text type. At test, participants judged the old-new status of 100 words (50 studied, 50 new distracters) presented in neutral text type and indicated the study text type (bold or italic) for any word judged to be old. The raw responses frequencies, aggregated across participants and items in each condition, are reported in the Appendix of Kuhlmann and Touron. We constructed a two-group version of the 2HT-MPT model of source monitoring displayed in **Figure 3**, yielding separate parameter estimates of the four parameters (D , d , b , and g) for the younger (subscript YA) and older (subscript OA) adults, and used multiTree (Moshagen, 2010) for parameter estimation and hypothesis tests. The data file, equation files, and multiTree file of this application example are provided as **Supplementary Material**.

The two-group joint standard model fit the data well, $G^2(4) = 6.75, p = 0.150$. **Table 1** presents parameter estimates for both age groups. Given the good fit of the model to the data, the parameter estimates may be interpreted and submitted to inferential tests. In this standard model, one can test the main effect of age

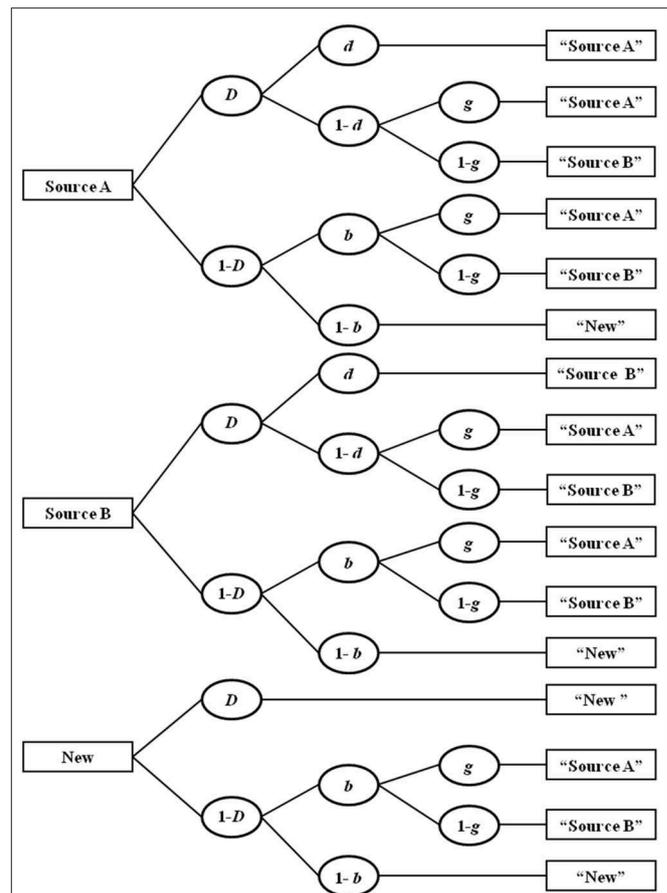


FIGURE 3 | Four-parameter version of the two-high-threshold multinomial processing tree model of source monitoring. D , probability of recognizing a previously presented item or detecting a distractor item; d , probability of remembering the source an item was presented in; b , probability of guessing that an unrecognized item is old; g , probability of guessing that an item was presented in Source A. Adapted from Bayen et al. (1996, p. 202).

group on item and source memory, respectively, via standard parameter comparisons. Restricting the item memory parameters to be equal across younger and older adults (i.e., $D_{YA} = D_{OA}$) significantly worsens model fit, $\Delta G^2(1) = 16.90, p < 0.001$, showing that, as to be expected, older adults’ item recognition is significantly poorer than younger adults’. Likewise, restricting the source-memory parameters to be equal across younger and older adults (i.e., $d_{YA} = d_{OA}$) significantly worsens model fit, $\Delta G^2(1) = 8.25, p = 0.004$. Again, older adults’ source memory is significantly poorer than younger adults’. Thus, these inferential tests confirm to-be-expected simple main effects of age group on both item and source memory. However, these tests do not assess whether the age effect is more pronounced on source than on item memory as predicted by the associative deficit theory.

In order to test the age group \times memory type interaction, we proceeded as detailed above. Specifically, we introduced two new shrinkage parameters, α_{AI} and α_{AS} , to reflect the effect of age group (A ; in the direction $OA < YA$) on item (I) and source (S) memory, respectively. Therefore, older adults’ item and source memory parameters are reparameterized as a proportion of

TABLE 1 | Parameter estimates from the original and the reparameterized four-parameter version of the two-high threshold multinomial model of source monitoring for younger and older adults in Kuhlmann and Touron (2011).

Sample	<i>D</i>	<i>d</i>	<i>b</i>	<i>g</i>	α_{AI}	α_{AS}	$G^2(4)$
Original model							
Younger	0.72 (0.01)	0.38 (0.03)	0.28 (0.02)	0.50 (0.02)	–	–	6.75
Older	0.64 (0.01)	0.25 (0.03)	0.31 (0.02)	0.54 (0.02)	–	–	
Reparameterized model							
Younger	0.72 (0.01)	0.38 (0.03)	0.28 (0.02)	0.50 (0.02)	–	–	6.75
Older	–	–	0.31 (0.02)	0.54 (0.02)	0.89 (0.02)	0.65 (0.10)	

D, probability of recognizing an old item or detecting a new distractor item; *d*, probability of remembering the source (i.e., text type) of a recognized old item; *b*, probability of guessing “old” if item is not recognized; *g*, probability of guessing “italic” if text type source is not remembered; α_{AI} , age-related decline in item recognition (relative to younger adults’ item recognition); α_{AS} , age-related decline in source memory (relative to younger adults’ source memory). $G^2(4)$ values smaller than 9.49 indicate a good model fit ($p > 0.05$).

younger adults’ respective memory parameter as follows (see **Supplementary Material** for full set of model equations):

$$D_{OA} = \alpha_{AI} \cdot D_{YA} \quad (15)$$

$$d_{OA} = \alpha_{AS} \cdot d_{YA} \quad (16)$$

This reparameterized model yielded the same model fit (G^2) as the original model, implying that the data satisfy the implemented same-direction order-constraints. In addition, none of the shrinkage parameter estimates converged to the boundary value 1 (or 0). We further ensured that identifiability holds for the reparameterized model, as demonstrated for the original model by Bayen et al. (1996), using multiTree’s tools for identifiability check: We confirmed local identifiability by repeatedly estimating the model parameters from the Kuhlmann and Touron (2011) data 1,000 times. No deviations occurred in any of the parameter estimates (i.e., all deviations ≤ 0.00001). Similarly, the simulated identifiability analysis based on 1,000 simulations with the same number of observations per tree as in the Kuhlmann and Touron data example, yielded only minor deviations between actual and recovered model parameters beyond the second decimal (maximum deviation of -0.00198 ; tolerance 0.00001) in less than 0.5% of the cases (i.e., 39 out of 1,000) and an average deviation < 0.00001 . These rare minor deviations can be explained by numerical inaccuracies. The output files from both identifiability checks are provided in **Supplementary Material**.

Parameter estimates from the reparameterized model are displayed in **Table 1**. Note that estimates for non-reparameterized parameters (i.e., *b*, *g*, D_{YA} , and d_{YA}) are identical to their estimates in the original model, whereas D_{OA} and d_{OA} have been replaced by the shrinkage parameters reflecting the ratios of the original memory parameter estimates for older relative to younger adults, that is $\hat{\alpha}_{AI} = \hat{D}_{OA}/\hat{D}_{YA} = 0.64/0.72 = 0.79$, and $\hat{\alpha}_{AS} = \hat{d}_{OA}/\hat{d}_{YA} = 0.25/0.38 = 0.65$. In essence, these values represent the simple main effect estimates of age (A) on item (I) and source (S) memory, respectively. Thus, there are decreases in item and source memory with increasing adult age, as indicated by the parameter estimates $\hat{\alpha}_{AI} < 1$ and $\hat{\alpha}_{AS} < 1$. The 95% CIs for $\hat{\alpha}_{AI}$, [0.85, 0.92], and for $\hat{\alpha}_{AS}$, [0.45, 0.85], exclude 1, suggesting significant age-related decreases in

item and source memory, in line with the earlier conducted simple main effect tests.

Most importantly, we can now directly compare the age-group effect on item vs. source memory by restricting $\alpha_{AI} = \alpha_{AS}$ to test the age group x memory type interaction postulated predicted by the associative deficit hypothesis. Descriptively, with $\hat{\alpha}_{AS} < \hat{\alpha}_{AI}$, there appears to be a larger age-group difference in source (associative) memory than in item memory, as implied by this theory, and restricting both α parameters to be equal indeed significantly worsens model fit, $\Delta G^2(1) = 4.07$, $p = 0.04$, indicating that the age difference is indeed significantly larger for source than for item memory—a strictly ordinal interaction as predicted by the associative deficit theory.

GENERALIZATION

So far, we have focused on two-way interactions. Interactions of higher orders (with three or more factors involved) imply that lower-level interactions (e.g., a two-way interaction) vary between the levels of another factor. The method of testing interactions as introduced here can be generalized to this scenario in a straightforward way. For example, we can introduce a third training factor to the experimental design illustrated in **Figure 1** such that both processes were assessed in both conditions in an untrained and in a trained group. That is, there are two sets of parameter estimates per condition one for untrained and one for trained, which can be denoted by a third subscript index of “1” for untrained and “2” for trained (i.e., θ_{111} is the estimate of process θ_1 in the untrained Condition 1 whereas θ_{112} is the estimate of process θ_1 in the trained Condition 1). As previously, we assume a main effect of condition with lower parameter estimates in Condition 2 than 1 and a main effect of process with lower probabilities of θ_2 than θ_1 . In addition, we now assume higher process estimates in the trained conditions for both Conditions 1 and 2. Of interest may now be to test whether this training benefit differs between Conditions 1 and 2 for one or both of the processes. That is, we are interested in testing a three-way interaction of condition, process, and training.

Given this research question, we can again make use of the reparameterization method introduced above. One way of

reparameterizing the model would be to introduce the training main effect α_T , quantifying the relative decrease in the untrained (u) compared to the trained (t) group in the first step. In a second step, the conditional (i.e., simple) main effects of condition within the two training groups, $\alpha_{C|u}$ and $\alpha_{C|t}$, respectively, would be specified. In the third and final step, one would introduce the second-order conditional main effects of process within the four condition \times training combinations ($\alpha_{p|C1,u}$, $\alpha_{p|C1,t}$, $\alpha_{p|C2,u}$, $\alpha_{p|C2,t}$, respectively). Applied simultaneously, all remaining parameters are reparameterized as a function of the largest parameter, θ_{112} , as follows:

$$\begin{aligned}\theta_{111} &= \alpha_T \cdot \theta_{112} \\ \theta_{121} &= \alpha_T \cdot \alpha_{C|u} \cdot \theta_{112} \\ \theta_{122} &= \alpha_{C|t} \cdot \theta_{112} \\ \theta_{211} &= \alpha_T \cdot \alpha_{p|C1,u} \cdot \theta_{112} \\ \theta_{212} &= \alpha_T \cdot \alpha_{p|C1,t} \cdot \theta_{112} \\ \theta_{221} &= \alpha_T \cdot \alpha_{C|u} \cdot \alpha_{p|C2,u} \cdot \theta_{112} \\ \theta_{222} &= \alpha_T \cdot \alpha_{C|t} \cdot \alpha_{p|C2,t} \cdot \theta_{112}\end{aligned}\quad (17)$$

Assuming that this reparameterized model fits the data as well as the model without order constraints, several (ordinal) interaction tests can be conducted. For example, if we would impose the equality constraints $\alpha_{C|u} = \alpha_{C|t}$ and $\alpha_{p|C1,u} = \alpha_{p|C1,t} = \alpha_{p|C2,u} = \alpha_{p|C2,t}$ simultaneously (i.e., four equality constraints in total), this constraint would test the hypothesis that there are no interactions whatsoever. In other words, for each of the three factors of the design, single decrease parameters (α_T , α_C , and α_p), representing the main effects of each factor, suffice to describe the relative differences between conditions, irrespective of the levels of other factors. If no specific prediction has been made before the study, this omnibus test of the H_0 of neither two-way nor three-way interactions is the recommended first test. The three-way interaction can then be tested using the second-order conditional main effects of process within the four condition \times training combinations (i.e., $\alpha_{p|C1,u}$, $\alpha_{p|C1,t}$, $\alpha_{p|C2,u}$, and $\alpha_{p|C2,t}$, respectively). For example, if $\alpha_{p|C1,u} = \alpha_{p|C1,t}$ but $\alpha_{p|C2,u} \neq \alpha_{p|C2,t}$ there would be a three-way interaction such that in Condition 2 the training effect differs by process whereas in Condition 1 it does not differ by process.

Importantly, note that the order in which the parametric order constraints are implemented determines the research questions that can be tested. If one was for example interested in comparing the training benefits for each Condition \times Process condition, one should have first implemented the condition and process effects and implemented the training effect as a second-order conditional main effect for each Condition \times Process combination. This order also determines which two-way interactions can be tested. The reparameterization order suggested first would allow to test for the condition \times training effect (i.e., $\alpha_{C|u} \neq \alpha_{C|t}$) whereas including the process effect before the training effect would allow to test for the Condition \times Process interaction (i.e., $\alpha_{p|C1} \neq \alpha_{p|C2}$). In sum, the optimal order in which parameters should be reparameterized depends on the specific research hypotheses of interest.

Using this approach, a multitude of interaction test options can be derived. In fact, all types of interactions as defined in Log-Linear Models for three-way contingency tables (Read and Cressie, 1988; Agresti, 2002) can be tested in this way, the single exception being the model allowing for all possible two-way interactions (condition by process, condition by training, and training by process) but not the three-way interaction of these factors. To our knowledge, there is no way to test this specific hypothesis within the MPT framework. Despite this limitation, the approach can be used to address a large number of scientifically relevant interaction hypotheses for multi-factorial designs.

DISCUSSION

In this article, we explain how to test interactions on MPT model parameters. Our method relies on parametric order constraints (Knapp and Batchelder, 2004) that represent a factor's conditional main effects at each level of another factor, thereby allowing for a direct test of condition main effects against each other to test for an interaction. Crucially, parametric order constraints capture a (within- or between-subjects) factor's simple main effects as relative proportions (i.e., parameter ratios from the factor's different levels) rather than as absolute parameter differences, thus an interaction in the LLM sense is tested. We argue that the LLM concept is appropriate for MPT models because of the restriction of model parameters (i.e., probabilities) to the interval $[0, 1]$. Given this constraint, scaling simple main effects as absolute parameter differences would artificially imply an interaction in the presence of strong main effects, a problem that disappears when simple main effects are represented by parameter ratios. We explained in detail how both ordinal and disordinal two-way interactions on MPT model parameters can be tested. An empirical application example from cognitive aging proves the usefulness of our method to various psychological research questions. Finally, we explain how our method generalizes to higher-order interactions involving three or more factors. Our method can be easily implemented within the freely available software multiTree (Moshagen, 2010), operating on Windows, Mac and Unix. Thereby, users benefit from all other options provided by multiTree, for example the option to conduct an a priori power analysis of the interaction tests.

Limitations

Our proposed method for testing interactions on MPT model parameters is flexible and can be easily adapted to any MPT model. Thus, it provides a powerful tool for psychological research. Nonetheless, there are a few limiting aspects that one should be aware of before implementing this method. The fact that the order of parameter restrictions determines which specific interactions can be assessed in more complex designs with three or more factors, may be perceived as a limitation of this method. However, the same applies to ANOVA models for which there are different options to follow up with simple conditional main effects analyses. As always, a good theoretical foundation of the research study should ensure that the order of implementing

parameter restrictions in complex designs does not become an exploratory playground.

A more serious limitation may be the fact that one type of interaction as defined in Log-Linear Models for three-way contingency tables (Read and Cressie, 1988; Agresti, 2002) cannot be tested with this method. More specifically, it is not possible to set up a model that simultaneously allows for all possible two-way interactions (Factor 1 \times Factor 2, Factor 1 \times Factor 3, Factor 2 \times Factor 3) but omits the three-way interaction of these factors (Factor 1 \times Factor 2 \times Factor 3). Nonetheless, all other types of interactions can be tested and as there are ways to specifically test each two-way interaction, we think this does not pose a real limitation to the applicability of MPT interaction tests for nearly all psychological research questions.

Finally, we have focused on MPT model applications based on data (i.e., observed frequencies) aggregated across items and participants within each experimental condition, as is traditionally done in MPT research (Batchelder and Riefer, 1999; Chechile, 2009; Erdfelder et al., 2009). A potentially problematic aspect of this “complete pooling” approach is that observations are treated as independently and identically distributed across the response categories. More recent approaches, mostly based on Bayesian methods and partial pooling, have been developed to estimate MPT parameters in the presence of heterogeneity among observations (Klauer, 2006, 2010; Stahl and Klauer, 2007; Smith and Batchelder, 2008, 2010; Matzke et al., 2015; Heck et al., 2018). Notably, as our approach to test interactions is based on reparameterized standard MPT models, it can also be used in combination with all of these hierarchical MPT approaches as long as no between-subjects factor is involved. When Bayesian hierarchical MPT models are used, however, carefully note recommendations regarding prior distributions for order-constrained parameters as standard settings may not be appropriate (Heck et al., 2015).

We recommend that users test whether heterogeneity is present in their frequency data (see Smith and Batchelder, 2008, for such tests). If significant heterogeneity is present, parameter estimates based on complete pooling should be compared with partial-pooling approaches to assess potential risks of aggregation bias in their data. Note that particularly in experimental MPT research, complete pooling of data may have advantages even if some moderate degree of heterogeneity is present in the data (e.g., Chechile, 2009).

REFERENCES

- Agresti, A. (2002). *Categorical Data Analysis, 2nd Edn.* New York, NY: Wiley.
- Batchelder, W. H., and Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychol. Rev.* 87, 375–397. doi: 10.1037/0033-295X.87.4.375
- Batchelder, W. H., and Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *Br. J. Math. Statist. Psychol.* 39, 129–149. doi: 10.1111/j.2044-8317.1986.tb00852.x

CONCLUSION

In summary, MPT models are valuable measurement tools for psychological research (cf., Erdfelder et al., 2009) but its previous applications have been primarily restricted to analyses of main effects only. With the method outlined in this article, various types of interactions can be tested on MPT model parameters, rendering this tool even more powerful for future psychological research.

DATA AVAILABILITY STATEMENT

The aggregated raw response frequency data are provided in the Appendix of Kuhlmann and Touron (2011). We additionally provide this data file as well as all our analysis files in the **Supplementary Material**. The individual response frequencies can be obtained from BK, kuhlmann@psychologie.uni-mannheim.de.

AUTHOR CONTRIBUTIONS

BK drafted the manuscript, provided the empirical data, and conducted the statistical analyses. All three authors contributed equally to the methodological aspects of this paper and improved and revised the manuscript conjointly.

FUNDING

BK and EE were partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—GRK 2277 Statistical Modeling in Psychology. BK was additionally supported by DFG grant KU 3329/1-1.

ACKNOWLEDGMENTS

We thank Dayna R. Touron for her contributions to the empirical data used for the example.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02364/full#supplementary-material>

Supplementary Material | Data and analyses files.

- Batchelder, W. H., and Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.* 6, 57–86. doi: 10.3758/BF03210812
- Bayen, U. J., Murnane, K., and Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 197–215. doi: 10.1037/0278-7393.22.1.197
- Bröder, A., and Meiser, T. (2007). Measuring source memory. *Z. Psychol.* 215, 52–60. doi: 10.1027/0044-3409.215.1.52
- Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *J. Math. Psychol.* 53, 562–576. doi: 10.1016/j.jmp.2009.06.005

- Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall.
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., and Nadarevic, L. (2009). Multinomial processing tree models: a review of the literature. *Z. Psychol.* 217, 108–124. doi: 10.1027/0044-3409.217.3.108
- Erdfelder, E., Cüpper, L., Auer, T.-S., and Undorf, M. (2007). The four-states model of memory retrieval experiences. *Z. Psychol.* 215, 61–71. doi: 10.1027/0044-3409.215.1.61
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Mem. Cog.* 16, 309–313. doi: 10.3758/BF03197041
- Heck, D. W., Arnold, N. R., and Arnold, D. (2018). TreeBUGS: an R package for hierarchical multinomial-processing-tree modeling. *Behavior Res. Methods* 50, 264–284. doi: 10.3758/s13428-017-0869-7
- Heck, D. W., Wagenmakers, E.-J., and Morey, R. D. (2015). Testing order constraints: qualitative differences between Bayes factors and normalized maximum likelihood. *Statist. Probab. Lett.* 105, 157–162. doi: 10.1016/j.spl.2015.06.014
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* 6, 65–70.
- Hu, X., and Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika* 59, 21–47. doi: 10.1007/BF02294263
- Hu, X., and Phillips, G. A. (1999). GPT.EXE: A powerful tool for the visualization and analysis of general processing tree models. *Behav. Res. Methods Instrum. Comput.* 31, 220–234. doi: 10.3758/BF03207714
- Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychol. Bull.* 114, 3–28. doi: 10.1037/0033-2909.114.1.3
- Klauer, K. C. (2006). Hierarchical multinomial processing tree models: a latent-class approach. *Psychometrika* 71, 7–31. doi: 10.1007/s11336-004-1188-3
- Klauer, K. C. (2010). Hierarchical multinomial processing tree models: a latent-trait approach. *Psychometrika* 75, 70–98. doi: 10.1007/S11336-009-9141-0
- Klauer, K. C., Musch, J., and Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychol. Rev.* 107, 852–884. doi: 10.1037/0033-295X.107.4.852
- Knapp, B. R., and Batchelder, W. H. (2004). Representing parametric order constraints in multi-trial applications of multinomial processing tree models. *J. Math. Psychol.* 48, 215–229. doi: 10.1016/j.jmp.2004.03.002
- Kuhlmann, B. G., and Touron, D. R. (2011). Older adults' use of metacognitive knowledge in source monitoring: spared monitoring but impaired control. *Psychol. Aging* 26, 143–149. doi: 10.1037/a0021055
- Matzke, D., Dolan, C. V., Batchelder, W. H., and Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika* 80, 205–235. doi: 10.1007/s11336-013-9374-9
- Meiser, T. (2014). Analyzing stochastic dependence of cognitive processes in multidimensional source recognition. *Exp. Psychol.* 61, 402–415. doi: 10.1027/1618-3169/a000261
- Moshagen, M. (2010). multiTree: a computer program for the analysis of multinomial processing tree models. *Behav. Res. Methods* 42, 42–54. doi: 10.3758/BRM.42.1.42
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: tests of an associative deficit hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1170–1187. doi: 10.1037//0278-7393.26.5.1170
- Read, T. R. C., and Cressie, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York, NY: Springer.
- Riefer, D. M., and Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychol. Rev.* 95, 318–339. doi: 10.1037/0033-295X.95.3.318
- Rummel, J., Boywitt, C. D., and Meiser, T. (2011). Assessing the validity of multinomial models using extraneous variables: an application to prospective memory. *Q. J. Exp. Psychol.* 64, 2194–2210. doi: 10.1080/17470218.2011.586708
- Singmann, H., and Kellen, D. (2013). MPTinR: analysis of multinomial processing tree models in R. *Behav. Res. Methods* 45, 560–575. doi: 10.3758/s13428-012-0259-0
- Smith, J. B., and Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychon. Bull. Rev.* 15, 713–731. doi: 10.3758/PBR.15.4.713
- Smith, J. B., and Batchelder, W. H. (2010). Beta-MPT: multinomial processing tree models for addressing individual differences. *J. Math. Psychol.* 54, 167–183. doi: 10.1016/j.jmp.2009.06.007
- Stahl, C., and Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behav. Res. Methods* 39, 267–273. doi: 10.3758/BF03193157

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Kuhlmann, Erdfelder and Moshagen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.