Running head: Order Constraints

Parametric Order Constraints in Multinomial Processing Tree Models:

An Extension of Knapp and Batchelder (2004)

Karl Christoph Klauer, Henrik Singmann, and David Kellen

Albert-Ludwigs-Universität Freiburg

Author Note

Karl Christoph Klauer, Institut für Psychologie; Henrik Singmann, Institut für
Psychologie; David Kellen, Institut für Psychologie.

The first two authors contributed to equal extents to the manuscript.

Correspondence concerning this article should be addressed to K. C. Klauer
at the Institut für Psychologie, Albert-Ludwigs-Universität Freiburg,
D-79085 Freiburg, Germany. Electronic mail may be sent to
christoph.klauer@psychologie.uni-freiburg.de.

Address information for Christoph Klauer (corresponding author):

- E-mail: christoph.klauer@psychologie.uni-freiburg.de

- Phone: +49 761 2032469

- Postal address:

  – Institut für Psychologie

  – Albert-Ludwigs-Universität Freiburg

  – D-79085 Freiburg

  – Germany

Abstract

Multinomial processing tree (MPT) models are tools for disentangling the contributions of latent cognitive processes in a given experimental paradigm. The present note analyzes MPT models subject to order constraints on subsets of its parameters. The constraints that we consider frequently arise in cases where the response categories are ordered in some sense such as in confidence-rating data, Likert scale data, where graded guessing tendencies or response biases are created via base-rate or payoff manipulations, in the analysis of contingency tables with order constraints, and in many other cases. We show how to construct an MPT model without order constraints that is statistically equivalent to the MPT model with order constraints. This new closure result extends the mathematical analysis of the MPT class, and it offers an approach to order-restricted inference that extends the approaches discussed by Knapp and Batchelder (2004). The usefulness of the method is illustrated by means of an analysis of an order-constrained version of the two-high-threshold model for confidence ratings.

KEYWORDS: Multinomial processing tree models, mathematical models, categorical data, multinomial distribution

Multinomial processing tree (MPT) models are used to measure cognitive processes in many areas of psychology (for reviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009). They are models for categorical data. MPT models are typically tailored to a given experimental paradigm and specify how the most important processes assumed to be involved in data generation in the paradigm interact to produce observable responses.

As an example consider the two-high-threshold model (2HTM; Snodgrass & Corwin, 1988). The 2HTM is tailored to memory experiments in which old/new judgments are requested for previously studied items intermixed with new items. In many such experiments, participants are also asked to rate their confidence in each "old" or "new" judgment. Figure 1 shows a version of the model for a confidence rating scale with three points, labeled "high", "medium", and "low" (Bröder, Kellen, Schütz, & Rohrmeier, 2013; Klauer & Kellen, 2011). Responses are mediated via three latent states, labeled "detect old", "detect new", and "no detection".

Parameters D_o and D_n define a stimulus-state mapping. D_o is the probability of entering the "detect old" state for an old item; D_n of entering the "detect new" state for a new item; the "no detection" state is entered with probability $1 - D_o$ and $1 - D_n$ for old and new items, respectively. The remaining parameters define state-response mappings. Given one of the two "detect" states, the old/new judgment is invariably correct as regards the old versus new status of the test item, and the parameters s_l, s_m, and s_h ($s_l + s_m + s_h = 1$) quantify the probabilities of selecting, in order, the low, medium, and high confidence level in the old/new response.[1] In the absence of detection, there is a guessing bias captured by probability parameter g, quantifying the probability of guessing "old" rather than "new". Given that "old" is guessed, parameters o_l, o_m, and o_h with $o_l + o_m + o_h = 1$ parameterize the probabilities for the three confidence levels; given that "new" is guessed, n_l, n_m, and n_h parameterize these probabilities.

As can be seen, MPT models assume that observed category counts arise

from processing branches consisting of separate conditional links or stages. Each branch probability is the product of its conditional link probabilities, and more than one branch can terminate in the same observed category (Hu & Batchelder, 1994).

In most cases, the models are eventually represented as so-called binary MPT models (Purdy & Batchelder, 2009), because many software tools for analyzing MPT models require binary MPT models as input. In a binary MPT model, exactly two links go out from each non-terminal node. The two links are labeled by two parameters that sum to one. One of these is redundant and is replaced by one minus the other parameter so that the remaining model parameters are functionally independent, each such parameter ranging from 0 to 1. It is straightforward to transform a non-binary MPT model into a statistically equivalent binary MPT model (Hu & Batchelder, 1994). Two models are statistically equivalent if they can predict the same sets of response probabilities.

In applications, it is not uncommon that order constraints are predicted to hold for subsets of the functionally independent parameters of binary MPT models (Baldi & Batchelder, 2003; Knapp & Batchelder, 2004), and Knapp and Batchelder have shown that the model class is closed under one or more non-overlapping linear orders of parametric constraints. That is, a new non-constrained binary MPT model can be constructed using a different set of functionally independent parameters that is statistically equivalent to the original model with the order constraints.

Here, we consider a different set of order constraints that regularly arise in applications and that are not covered by Knapp and Batchelder (2004). The order constraints frequently arise where response categories are ordered in some sense such as for confidence ratings or Likert scales. They also arise where participants discriminate between two or more categories of items and the probabilities of guessing the categories in "no detection" states can be assumed to be ordered; for example, because base rates or payoffs systematically differ between the categories. Our results also apply to the important case of order constraints on the probabilities of a

multinomial or product-multinomial distribution that is frequently encountered within and outside psychology (e.g., Agresti & Coull, 2002). We show that an MPT model with the order constraints can be represented in the form of a statistically equivalent non-constrained MPT model. This new closure property contributes to the structural analysis of the MPT class and is immediately useful for analyzing cases in which the order constraints are to be imposed upon the parameters as exemplified below.

Considering, for example, the 2HTM for confidence ratings, a psychologically plausible constraint on the parameters for the confidence levels in the "no detection" state is that the preference for a given confidence level should decline from lowest to highest confidence levels, reflecting the respondent's uncertainty in the absence of detection: $o_l \geq o_m \geq o_h$ and $n_l \geq n_m \geq n_h$. Conversely, in "detect" states, the preference for a given confidence level should increase from lowest to highest confidence, at least for scales with only a few confidence levels: $s_l \leq s_m \leq s_h$.

Imposing such constraints sharpens the distinctions between "no detection" and "detect" states by highlighting plausible qualitative differences between them. When satisfied by the underlying probability distribution, the constraints contribute to making the estimation of the parameters D_n and D_o of the stimulus-state mapping more precise, focused, and robust, and they considerably increase the model's parsimony as elaborated on below.

These order constraints are imposed on functionally dependent parameters (e.g., s_l, s_m, and s_h have to sum to 1 and are therefore not independent). Hence, they are not covered by Knapp and Batchelder's (2004) approach to order constraints for independent parameters. Nevertheless, it is possible to express them in the language of MPT models.

The next section describes how to transform MPT models with order constraints of this kind into equivalent non-constrained MPT models.[2] Finally, we illustrate the new method by comparing versions of the 2HTM with and without order constraints in terms of model complexity and in terms of their description of a dataset

by Koen and Yonelinas (2010). The general discussion expands on the advantages of the new method for estimation and inference with order-constrained models.

Order Constraints on Multinomial Probabilities

We consider a basic subtree with a root, no other non-terminal node, and two or more links going out from its root as shown on the left side of Figure 2. This subtree might occur at one or more places in the tree representation of the complete model. With regard to the overall MPT model, its terminal nodes $A_1, \ldots, A_k$ represent either other subtrees or observable categories. For example, the subtree with three links labeled by parameters s_l, s_m, and s_h in the above 2HTM occurs at two places in the processing tree representation (see Figure 1).

Our basic result describes how to represent a linear order $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_k$ on the parameters of the subtree by means of a statistically equivalent MPT model without order constraints on the parameters. As shown in Figure 2 the solution is to replace each occurrence of the subtree in question by the subtree on the right side in Figure 2. Replacement means that the tree on the right side replaces the subtree on the left side wherever it occurs in the processing-tree representation. Furthermore, whatever is appended at the terminal node A_j of an occurrence of the original subtree is appended in the replacing subtree wherever a terminal node labeled A_j occurs. As for the reparameterizations in Knapp and Batchelder (2004) this regularly implies an increase in the size of the processing tree.

We prove the following two theorems:

Theorem 1: For the tree shown on the right side of Figure 2 and any set of non-negative parameters λ_i, $i = 1, \ldots, k$, with $\sum_{i=1}^{k} \lambda_i = 1$, the probabilities of outcomes A_i are ordered as $P(A_1) \geq P(A_2) \geq \ldots \geq P(A_k)$.

Theorem 1 states that the tree indeed imposes an order constraint on its k outcome probabilities. Theorem 2 complements this by showing that any ordered set of probabilities can be represented in this form:

Theorem 2: For any set of ordered probabilities η_i with $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_k$,

$\sum_i \eta_i = 1$, there exist non-negative values λ_i, $i = 1, \ldots, k$ with $\sum_i \lambda_i = 1$ such that the outcome probabilities of the tree on the right side of Figure 2 are given by $P(A_i) = \eta_i$, $i = 1, \ldots, k$.

Proof of Theorem 1. Note that the outcome probabilities of the tree on the right side of Figure 2 are mixtures with mixture coefficients λ_j. The first mixture component is given by the probability distribution with $P(A_1) = 1$ and $P(A_j) = 0$, $j > 1$; the second by $P(A_1) = P(A_2) = 1/2$ and $P(A_j) = 0$, $j > 2$; the last by $P(A_j) = 1/k$ for all j. Each mixture component satisfies the inequalities, $P(A_j) \geq P(A_{j+1})$, $j = 1, \ldots, k - 1$. This completes the proof.

Proof of Theorem 2. Define mixture coefficients λ_j as follows:

$$\lambda_k = k\eta_k \text{ and } \lambda_j = j(\eta_j - \eta_{j+1}), j = 1, \ldots, k - 1. \tag{1}$$

By the premises of Theorem 2, it is immediate that $\lambda_j \geq 0$. It is furthermore easy to see via simple manipulations that $\sum_{j=1}^{k} \lambda_j = \sum_{j=1}^{k} \eta_j = 1$ and that $\eta_j = \sum_{i=j}^{k} \frac{\lambda_i}{i}$. Hence, $P(A_j) = \sum_{i=j}^{k} \frac{\lambda_i}{i} = \eta_j$, $j = k, k - 1, \ldots, 1$. This completes the proof.

Extensions

In this section, we consider linear-order constraints on only a subset of the η_i, non-overlapping linear-order constraints on the η_i, more general partial orders, and all such orders for $k = 3$ and $k = 4$.

Linear orders on subsets and multiple non-overlapping linear orders. Theorems 1 and 2 cover the case of a total order on all k parameters η_i. It is straightforward to extend the results to the case of a partial linear order on only a subset of the η_i. For example, assume $k = 6$ and that the constraint $\eta_1 \geq \eta_2 \geq \eta_3$ is to be imposed. It is easy to see that this translates into the constraint $\theta_1 \geq \theta_2 \geq \theta_3$ in the reparameterization depicted in Figure 3, where $\theta_1 + \theta_2 + \theta_3 = 1$ so that Theorems 1 and 2 apply.

The reparameterization in Figure 3 also shows how two or more non-overlapping linear-order constraints can be imposed upon the η_i with $\sum_i \eta_i = 1$. For example, if in addition $\eta_4 \geq \eta_5 \geq \eta_6$ is to be imposed, this could be implemented

by imposing the constraint $\theta_4 \geq \theta_5 \geq \theta_6$, where $\theta_4 + \theta_5 + \theta_6 = 1$, in terms of the parameters θ_i in Figure 3.

General partial orders. More general partial orders can often be treated by the following idea. The probability distributions $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$ that satisfy a set of linear inequality constraints form a convex polytope. Specifically, they can exhaustively be represented as mixtures of certain fixed probability distributions $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_l$ that we refer to as vertices. For small problems, the vertices can be found graphically; in complex cases, linear programming algorithms can be used.

For example, for the linear order with $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_k$, the vertices are given by the above-mentioned k mixture components, $\boldsymbol{\eta}_1 = (1, 0, \ldots, 0)$, $\boldsymbol{\eta}_2 = (1/2, 1/2, 0, \ldots, 0)$, ..., $\boldsymbol{\eta}_k = (1/k, \ldots, 1/k)$, and the model parameters λ_j of the non-constrained MPT model expressing the order constraint are simply the mixture weights. The subtree following λ_j codes the probability distribution specified in vertex $\boldsymbol{\eta}_j$

Partial orders with $k = 3$. This immediately solves the case of orderings among k parameters that can be represented by k vertices. Consider, for example, $k = 3$, and the ordering $\eta_1 \leq \eta_2$ and $\eta_1 \leq \eta_3$. This defines a convex polytope with the three vertices $\boldsymbol{\eta}_1 = (0, 0, 1)$, $\boldsymbol{\eta}_2 = (0, 1, 0)$, and $\boldsymbol{\eta}_3 = (1/3, 1/3, 1/3)$. Hence, a non-constrained MPT model representing the order restrictions has three mixture coefficients λ_i, $i = 1, \ldots, 3$, as parameters. Each λ_i is linked to a subtree coding the respective probability distributions $\boldsymbol{\eta}_i$.

In other cases, more than k vertices are required. For example, the ordering $\eta_1 \geq \eta_2$, $\eta_1 \geq \eta_3$ defines a polytope with four vertices, $\boldsymbol{\eta}_1 = (1, 0, 0)$, $\boldsymbol{\eta}_2 = (1/2, 1/2, 0)$, $\boldsymbol{\eta}_3 = (1/2, 0, 1/2)$, and $\boldsymbol{\eta}_4 = (1/3, 1/3, 1/3)$. Using four mixture coefficients λ_i to represent the order-constrained MPT model as a non-constrained MPT model is possible, but results in an overparameterized model. This means that certain analyses and inferences available for MPT models without overparameterization cannot be done. For example, although it is possible to determine the model's maximum

log-likelihood and G^2 goodness-of-fit statistic and to bootstrap its distribution (e.g., Singmann & Kellen, 2013), it will not be possible to determine unique parameter estimates.

For the present example, it can however be shown that representing the four mixture coefficients λ_i by two independent parameters θ_1 and θ_2 with $0 \leq \theta_i \leq 1$ such that $\lambda_1 = (1 - \theta_1)(1 - \theta_2)$, $\lambda_2 = \theta_1(1 - \theta_2)$, $\lambda_3 = (1 - \theta_1)\theta_2$, and $\lambda_4 = \theta_1\theta_2$ is sufficient to span the entire polytope defined by the above vertices (see appendix). This immediately leads to a non-redundant parameterization using two independent parameters θ_1 and θ_2.

But an analogous construction does not in general guarantee this in other cases. Nevertheless, for most purposes it is sufficient that a non-redundant parameterization is found that covers the point of maximum likelihood (as determined, for example, via estimating the overparameterized model in a first step) and its local environment. This is usually not difficult to achieve departing from the vertex representation.

Partial orders with $k = 4$. Tables 1 and 2 show the vertex representations of the possible partial orders involving four outcomes A, B, C, and D as shown in Figure 4. The online supplemental material sketches a heuristic for determining non-redundant parameterizations of the mixture coefficients in these and more complex cases.

Application: Order Constraints on the 2HTM

As already noted, the 2HTM distinguishes "detect" states and a "no detection" state. These are latent states, and a state-response mapping is required to link the states to the observable responses. Klauer and Kellen (2011) and Bröder et al. (2013) proposed relatively unrestricted state-response mapping in which it is possible, for example, that extreme confidence ratings would be preferred in a "no detection" state and that low confidence ratings would be preferred in a "detect" state. This has prompted criticisms to the effect that the models deal with rating

scales in an arbitrary and post-hoc manner (e.g., Dube, Rotello, & Pazzaglia, 2013, Pazziaglia, Dube, & Rotello, 2013; see also Batchelder & Alexander, 2013). According to these critics, the model is overly complex (Dube, Rotello, & Heit, 2011).

Using the present results, it is possible to define an order-constrained 2HTM for confidence ratings, $2HTM_r$, that maps the "no detection" state so that cautious low confidence ratings receive most probability mass and more extreme confidence ratings successively less mass. Conversely, for detect states, the preference for confidence levels increases from low confidence to high confidence levels in $2HTM_r$. Specifically, $2HTM_r$ imposes the constraints that $s_h \geq s_m \geq s_l$ for the mapping from "detect" states to responses, and that $o_l \geq o_m \geq o_h$ and $n_l \geq n_m \geq n_o$ for the mapping of the "no detection" state to responses (see Figure 1 for the 2HTM). These constraints remove most of the less plausible state-response mappings that are admissible under the original 2HTM for confidence ratings. At the same time, they strongly curtail the mathematical flexibility of the model.

Because these constraints can now be implemented within the MPT framework, we can use standard MPT software to estimate, test, and analyze the model. For example, we used the R package MPTinR (Singmann & Kellen, 2013) to quantify model flexibility of the 2HTM and the $2HTM_r$ based on the minimum-description-length principle (Grünwald, 2007) that takes flexibility due to a model's functional form into account, including constraints on flexibility due to inequality constraints. In the present context, one representation of the minimum-description-length principle is given by the Fisher information approximation (FIA) index (but see Heck, Moshagen, & Erdfelder, 2014). FIA is a model selection index and describes model flexibility by a penalty that is added to a likelihood measure of the model's (mis)fit. The model with the smallest index value is preferred. Wu, Myung, and Batchelder (2010a, 2010b) developed methods to compute FIA for binary MPT models that is implemented in MPTinR.

The penalty term for model flexibility in FIA comprises two additive terms.

One term depends upon the number of parameters and the sample size similar to the penalty term in BIC. A second term quantifies model flexibility due to functional form, and 2HTM and 2HTM$_r$ differ in the size of this penalty term. Using the above reparameterization and MPTinR, we find the penalties due to flexibility to be -0.01 and -5.22 for 2HTM and 2HTM$_r$, respectively. Because FIA and the normalized maximum-likelihood index operate on a log-likelihood scale, this means that the loss in goodness of fit in terms of G^2 (Batchelder & Riefer, 1999) must be larger than $10.42 = 2[-0.01 - (-5.22)]$ before the parsimonious 2HTM$_r$ should be abandoned in favor of the non-constrained 2HTM. This demonstrates that the constraints on flexibility imposed by the order constraints are substantial.

For example, reanalyzing data from Koen and Yonelinas (2010, pure condition), the 2HTM without order constraints achieves a G^2 statistic of 3.35 and a FIA index of 8.28 (up to an additive constant). The 2HTM with order constraints achieves a G^2 statistic of 9.95 and a FIA index of 6.37. Thus, the loss in goodness of fit is outweighed by the parsimony of the constrained model, and the 2HTM$_r$ should be preferred.

Discussion

In this note, we extended Knapp and Batchelder's (2004) approach to order constraints in MPT models to the case of order restrictions on non-independent parameters constrained to sum to one. These restrictions frequently arise in cases where response outcomes themselves are ordered in some sense such as in confidence-rating data, Likert scale data (Böckenholt, 2012; Klauer & Kellen, 2011) or where graded guessing tendencies or response biases are created via base-rate or payoff manipulations. Alternatively, the restrictions can directly arise from theoretical predictions (e.g., Ragni, Singmann, & Steinlein, in press).

The results are useful because they make available the growing toolbox for statistical analyses of MPT model for the analysis of order-constrained MPT models. As already exemplified, the toolbox comprises algorithms and software for the

computation of FIA (Moshagen, 2010; Singmann & Kellen, 2013; Wu et al., 2010a, 2010b), but also a number of software tools to estimate and fit the models (see Klauer, Stahl, & Voss, 2012 for a review), algorithms for Bayesian hierarchical model extensions that capitalize on the model structure (Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, in press; Smith & Batchelder, 2010), algorithms and software for hierarchical latent-class extensions of MPT models in a classical inferential framework (Klauer, 2006, Stahl & Klauer, 2007), and algorithms for computing Bayes factors between competing MPT models (Vandekerckhove, Matzke, & Wagenmakers, in press).

The present results also apply to the important case of order constraints on the probabilities of an observable multinomial or product-multinomial distribution, a case with many occurrences within and outside psychology — for example, in the analysis of contingency tables with order constraints (Agresti & Coull, 2002). The case is treated by Robertson, Wright, and Dykstra (1988, chap. 5) via their elegant isotonic regression method. Note, however, that expressing order structures on observable category probabilities in the MPT framework has the advantage of making available the above-mentioned tools for estimation and inference in classical and Bayesian frameworks. As one example, using MPTinR (Singmann & Kellen, 2013), the distribution of goodness-of-fit statistic G^2 can be assessed via bootstrap methods under the null hypothesis that the constraints are truly in force, side-stepping the numerically difficult task to evaluate its asymptotic so-called $\bar{\chi}^2$ distribution.

On the theoretical side, these results contribute to the study of the class of MPT models. They show that the class is closed under this further set of order constraints. This flexibility is surprising given that most other classes of statistical models that we are aware of would not be invariant under transformations as in Equation 1, nor capable of expressing distributions with and without the present kind of order constraints within the same class of models. The ability of the model class to encompass the present and other kinds of meaningful order constraints adds to its

usefulness.

References

Agresti, A., & Coull, B. A. (2002). The analysis of contingency tables under inequality constraints. *Journal of Statistical Planning and Inference*, *107*, 45–73. doi: 10.1016/S0378-3758(02)00243-4

Baldi, P., & Batchelder, W. H. (2003). Bounds on variances of estimators for multinomial processing tree models. *Journal of Mathematical Psychology*, *47*, 467–470. doi: 10.1016/S0022-2496(03)00036-1

Batchelder, W. H., & Alexander, G. E. (2013). Discrete-state models: Comment on Pazzaglia, Dube, and Rotello (2013). *Psychological Bulletin*, *139*, 1204–1212. doi: 10.1037/a0033894

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review*, *6*, 57-86. doi: 10.3758/BF03210812

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665–678. doi: 10.1037/a0028111

Bröder, A., Kellen, D., Schütz, J., & Rohrmeier, C. (2013). Validating a two-high-threshold measurement model for confidence rating data in recognition. *Memory*, *21*, 916–944. doi: 10.1080/09658211.2013.767348

Dube, C., Rotello, C. M., & Heit, E. (2011). The belief bias effect is aptly named: A reply to Klauer and Kellen (2011). *Psychological Review*, *118*, 155-163. doi: 10.1037/a0019634

Dubé, C., Rotello, C. M., & Pazzaglia, A. M. (2013). The statistical accuracy and theoretical status of discrete-state MPT models: Reply to Batchelder and Alexander (2013). *Psychological Bulletin*, *139*, 1213–1220. doi: 10.1037/a0034453

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Assfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology*, *217*, 108-124. doi: 10.1027/0044-3409.217.3.108

Grünwald, P. (2007). *The minimum description length principle.* Cambridge, Mass.: MIT Press.

Heck, D. W., Moshagen, M., & Erdfelder, E. (2014). Model selection by minimum description length: Lower-bound sample sizes for the Fisher information approximation. *Journal of Mathematical Psychology*, *60*, 29–34. doi:

10.1016/j.jmp.2014.06.002

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*, 21-47. doi: 10.1007/BF02294263

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., & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review*, *17*, 465-478. doi: 10.3758/PBR.17.4.465

Klauer, K. C., & Kellen, D. (2011). Assessing the belief bias effect with ROCs: Reply to Dube, Rotello, and Heit (2010). *Psychological Review*, *118*, 164-173. doi: 10.1037/a0020698

Klauer, K. C., Stahl, C., & Voss, A. (2012). Multinomial models and diffusion models. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology. Abridged edition* (p. 331-354). New York: Guilford.

Knapp, B. R., & Batchelder, W. H. (2004). Representing parametric order constraints in multi-trial applications of multinomial processing tree models. *Journal of Mathematical Psychology*, *48*, 215–229. doi: 10.1016/j.jmp.2004.03.002

Koen, J. D., & Yonelinas, A. P. (2010). Memory variability is due to the contribution of recollection and familiarity, not to encoding variability. *Journal of Experimental Psychology. Learning, Memory and Cognition*, *36*, 1536-1542. doi: 10.1037/a0020448

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (in press). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42–54. doi: 10.3758/BRM.42.1.42

Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*. doi: 10.1037/a0033044

Purdy, B. P., & Batchelder, W. H. (2009). A context-free language for binary multinomial processing tree models. *Journal of Mathematical Psychology*, *53*, 547–561. doi: 10.1016/j.jmp.2009.07.009

Ragni, M., Singmann, H., & Steinlein, M. (in press). Theory comparison for generalized quantifiers. In P. Bello, M. Guarini, M. McShane, & B. Scassellatti (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Rao, C. R. (1973). *Linear statistical inference and its applications.* New York: Wiley. doi: 10.1002/9780470316436

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models with R. *Behavior Research Methods*, *45*, 560-575. doi: 10.3758/s13428-012-0259-0

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*, 167-183. doi: 10.1016/j.jmp.2009.06.007

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50. doi: 10.1037/0096-3445.117.1.34

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, *39*, 267-273. doi: 10.3758/BF03193157

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (in press). Model comparison and the principle of parsimony. In J. Busemeyer, J. Townsend, Z. J. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology.* Oxford: Oxford University Press.

Wu, H., Myung, J. I., & Batchelder, W. H. (2010a). Minimum description length model selection of multinomial processing tree models. *Psychonomic Bulletin & Review*, *17*, 275-286. doi: 10.3758/PBR.17.3.275

Wu, H., Myung, J. I., & Batchelder, W. H. (2010b). On the minimum description length complexity of multinomial processing tree models. *Journal of Mathematical Psychology*, *54*, 291 - 303. doi: DOI: 10.1016/j.jmp.2010.02.001

Appendix

We show that each probability distribution on three outcomes with $\eta_1 \geq \eta_2$ and $\eta_1 \geq \eta_3$ and $\sum_{i=1}^{3} \eta_i = 1$ can be represented as a mixture with independent parameters θ_1, θ_2, $0 \leq \theta_i \leq 1$ as follows:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = (1-\theta_1)(1-\theta_2)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \theta_1(1-\theta_2)\begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} + (1-\theta_1)\theta_2\begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix} + \theta_1\theta_2\begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}.$$

The proof proceeds by showing that this equation can be solved for θ_1 and θ_2 given η_1, η_2, and η_3. Note that $\eta_1 - \eta_2 - \eta_3 = (1 - \theta_1)(1 - \theta_2) - \frac{1}{3}\theta_1\theta_2$. Setting $\Delta = \eta_3 - \eta_2$, it follows that $\Delta = \frac{1}{2}(\theta_2 - \theta_1)$, hence $\theta_2 = 2\Delta + \theta_1$. Substituting this in the previous equation yields: $2\eta_1 - 1 = \eta_1 - \eta_2 - \eta_3 = \frac{1}{3}(3 - 6\Delta - 2\theta_1(3 - 2\Delta) + 2\theta_1^2)$. Some manipulations show that this is equivalent to

$$[\theta_1 - \tfrac{1}{2}(3 - 2\Delta)]^2 = [\tfrac{1}{2}(3 - 2\Delta)]^2 + 3(\eta_1 + \Delta - 1).$$

For this quadratic equation to be solvable in θ_1, it needs to be shown that $[\frac{1}{2}(3 - 2\Delta)]^2 + 3(\eta_1 + \Delta - 1) \geq 0$. This is equivalent to $[\frac{1}{2} + (1 - \Delta)]^2 \geq 3(1 - \eta_1 - \Delta)$. Because $\eta_1 + \eta_2 + \eta_3 = 1$, this is equivalent to $[\frac{1}{2} + \eta_1 + 2\eta_2]^2 \geq 6\eta_2$. Because $\eta_1 \geq \eta_2$, this is true if $[\frac{1}{2} + 3\eta_2]^2 \geq 6\eta_2$. This is equivalent to $(3\eta_2 - \frac{1}{2})^2 \geq 0$.

Hence, one solution of the above equation is $\theta_1 = \frac{1}{2}(3 - 2\Delta) - \sqrt{[\frac{1}{2}(3 - 2\Delta)]^2 - 6\eta_2}$. We will show that $0 \leq \theta_1 \leq 1$. $\theta_1 \geq 0$ is to see noting that $\frac{1}{2}(3 - 2\Delta) = \frac{1}{2} + \eta_1 + 2\eta_2 \geq 0$. Furthermore, because of this, $\theta_1 \leq 1$ is equivalent to $\frac{1}{2} + \eta_1 + 2\eta_2 - \sqrt{(\frac{1}{2} + \eta_1 + 2\eta_2)^2 - 6\eta_2} \leq 1$ or to $\eta_1 + 2\eta_2 - \frac{1}{2} \leq \sqrt{(\frac{1}{2} + \eta_1 + 2\eta_2)^2 - 6\eta_2}$. This is trivially true if the term to the left, $\eta_1 + 2\eta_2 - \frac{1}{2}$, is smaller than zero. If it is non-negative, on the other hand, this is equivalent to $(\eta_1 + 2\eta_2 - \frac{1}{2})^2 \leq (\frac{1}{2} + \eta_1 + 2\eta_2)^2 - 6\eta_2$, which is equivalent to $\eta_1 \geq \eta_2$ as simple manipulations show. Because the equations are symmetrical in θ_1 and θ_2, interchanging the roles of θ_1 and θ_2 and those of η_2 and η_3 shows that $\theta_2 = \theta_1 + 2\Delta$ also ranges between 0 and 1. This completes the proof.

Footnotes

[1]There are well-documented response-style effects such as preferring moderate over extreme responses or vice versa as moderated by contextual and personality factors (Böckenholt, 2012). In the light of these effects, it is reasonable to assume that "detect" states are not necessarily always mapped on the highest confidence level. For reasons of parsimony and model identifiability, we assume in the present case that the state-response mapping of confidence ratings for "detect old" and "detect new" states is the same (but see Klauer & Kellen, 2010).

[2]The non-constrained MPT models can be transformed into equivalent binary MPT models in a second step which we do not describe, because it is well known (Hu & Batchelder, 1994). Furthermore, given the maximum likelihood estimates of the parameters of the binary MPT model and an estimate of its Fisher information, maximum likelihood estimates of the parameters of the non-constrained MPT model as well as of the parameters of the original order-constrained MPT model and of their Fisher information matrices (for confidence intervals) can be obtained via standard methods (Rao, 1973, chap. 6a) using the first derivatives of the respective parameter transformation functions that transform these models' parameters into each other although there is as of yet no user-friendly software to accomplish this.

Table 1

Vertices for the Patterns in Figure 4

Pattern	Category	Vertices							
		1	2	3	4	5	6	7	8
I	A	1	1/2	1/3	1/4				
	B	0	1/2	1/3	1/4				
	C	0	0	1/3	1/4				
	D	0	0	0	1/4				
II	A	1	0	0	1/4				
	B	0	1	0	1/4				
	C	0	0	1	1/4				
	D	0	0	0	1/4				
III	A	1	0	1/3	1/4				
	B	0	1	1/3	1/4				
	C	0	0	1/3	1/4				
	D	0	0	0	1/4				
IV	A	1	0	1/2	1/4				
	B	0	1	1/2	1/4				
	C	0	0	0	1/4				
	D	0	0	0	1/4				
V	A	1	1/2	1/2	1/3	1/4			
	B	0	1/2	0	1/3	1/4			
	C	0	0	1/2	1/3	1/4			
	D	0	0	0	0	1/4			
VI	A	1	1/2	1/3	1/3	1/4			
	B	0	1/2	1/3	1/3	1/4			
	C	0	0	1/3	0	1/4			
	D	0	0	0	1/3	1/4			
VII	A	1	0	1/3	1/3	1/4			
	B	0	1	1/3	1/3	1/4			
	C	0	0	1/3	0	1/4			
	D	0	0	0	1/3	1/4			
VIII	A	1	0	1/3	0	1/4			
	B	0	1	1/3	1/2	1/4			
	C	0	0	1/3	0	1/4			
	D	0	0	0	1/2	1/4			
IX	A	1	1/2	1/2	1/3	1/3	1/4		
	B	0	1/2	0	1/3	1/3	1/4		
	C	0	0	1/2	1/3	0	1/4		
	D	0	0	0	0	1/3	1/4		
X	A	1	1/2	1/2	1/2	1/3	1/3	1/3	1/4
	B	0	1/2	0	0	1/3	1/3	0	1/4
	C	0	0	1/2	0	1/3	0	1/3	1/4
	D	0	0	0	1/2	0	1/3	1/3	1/4

Table 2

Non-Redundant Parameterizations of Mixture Weights for the Patterns in Figure 4

Ver-tices	Patterns				
	V	VI	VIII	IX	X
1	$(1-\theta_1)(1-\theta_2)(1-\theta_3)$	θ_1	θ_1	$(1-\theta_1)(1-\theta_2)(1-\theta_3)$	$(1-\theta_1)(1-\theta_2)(1-\theta_3)$
2	$\theta_1(1-\theta_2)(1-\theta_3)$	$(1-\theta_1)(1-\theta_2)(1-\theta_3)$	$(1-\theta_1)(1-\theta_2)(1-\theta_3)$	$\theta_1(1-\theta_2)(1-\theta_3)$	$\theta_1(1-\theta_2)(1-\theta_3)$
3	$(1-\theta_1)\theta_2(1-\theta_3)$	$(1-\theta_1)\theta_2(1-\theta_3)$	$(1-\theta_1)\theta_2(1-\theta_3)$	$(1-\theta_1)\theta_2(1-\theta_3)$	$(1-\theta_1)\theta_2(1-\theta_3)$
4	$\theta_1\theta_2(1-\theta_3)$	$(1-\theta_1)(1-\theta_2)\theta_3$	$(1-\theta_1)(1-\theta_2)\theta_3$	$\theta_1\theta_2(1-\theta_3)$	$(1-\theta_1)(1-\theta_2)\theta_3$
5	θ_3	$(1-\theta_1)\theta_2\theta_3$	$(1-\theta_1)\theta_2\theta_3$	$(1-\theta_2)\theta_3$	$\theta_1\theta_2(1-\theta_3)$
6				$\theta_2\theta_3$	$\theta_1(1-\theta_2)\theta_3$
7					$(1-\theta_1)\theta_2\theta_3$
8					$\theta_1\theta_2\theta_3$

Note. Patterns I to IV employ four mixture coefficients which are trivial to parameterize with three non-redundant parameters θ_i with $0 \leq \theta_i \leq 1$, $i = 1, \ldots, 3$. We believe that no parameterization exists for pattern VII that exhaustively represents all probability distributions with the order constraints using three non-redundant parameters, but did not find a proof for the non-existence of such a parameterization. We used numerical methods to ascertain that the parameterizations shown exhaust the space of probability distributions with the appropriate order constraints for all practical purposes (see online supplemental materials).

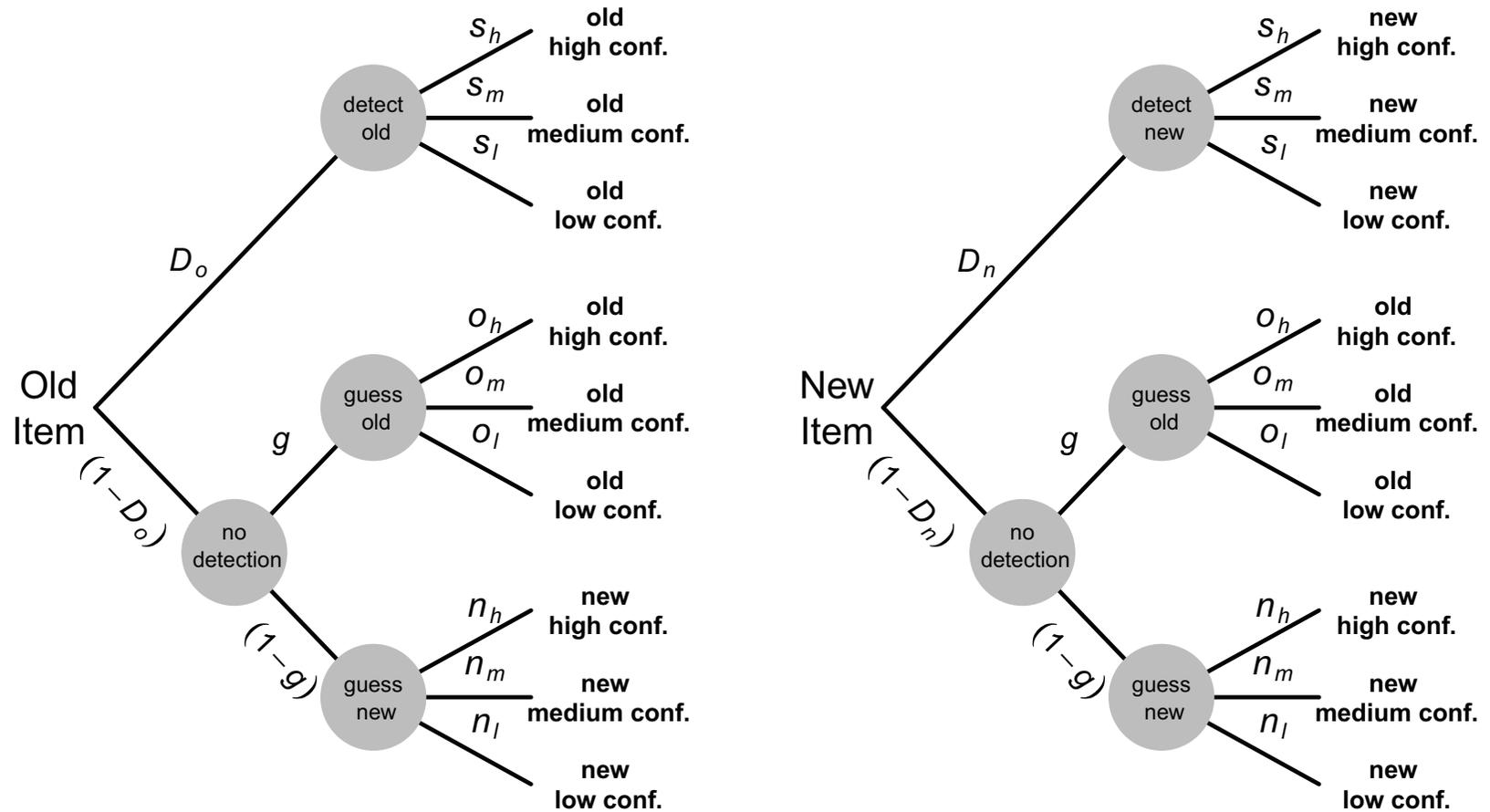# Two−High−Threshold Model (2HTM)

Figure 1. The 2HTM model for confidence ratings. Response categories are shown in the middle; the tree to the left models processing for old items; the tree to the right for new items.
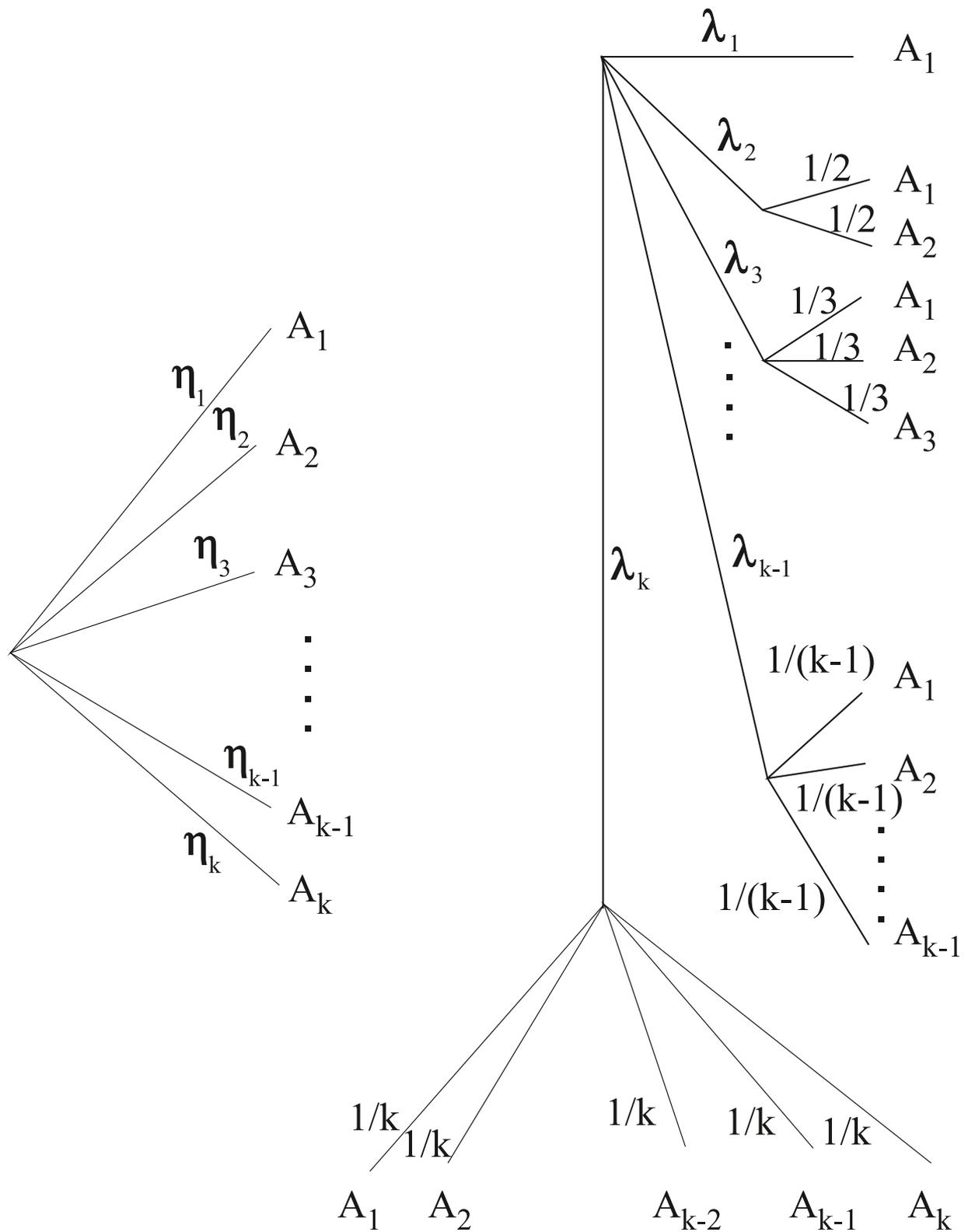
Figure 2. The tree to the right is a statistically equivalent reparameterization of the tree to the left for ordered parameters $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_{k-1} \geq \eta_k$.
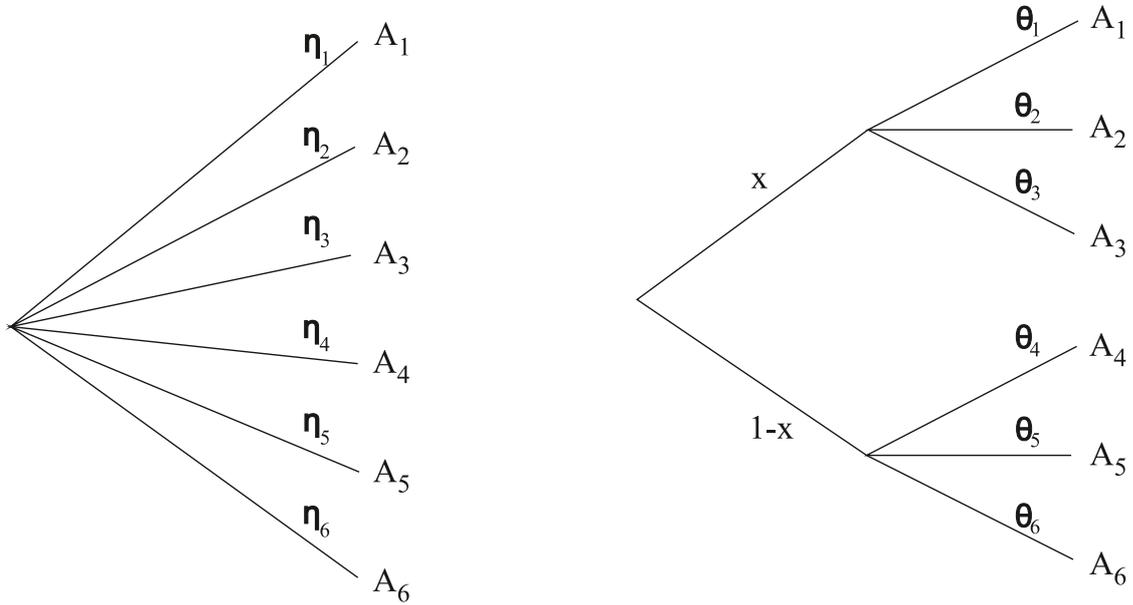
Figure 3. The tree to the right is a statistically equivalent reparameterization of the tree to the left, if $\sum_i \eta_i = \sum_{i=1}^3 \theta_i = \sum_{i=4}^6 \theta_i = 1$.
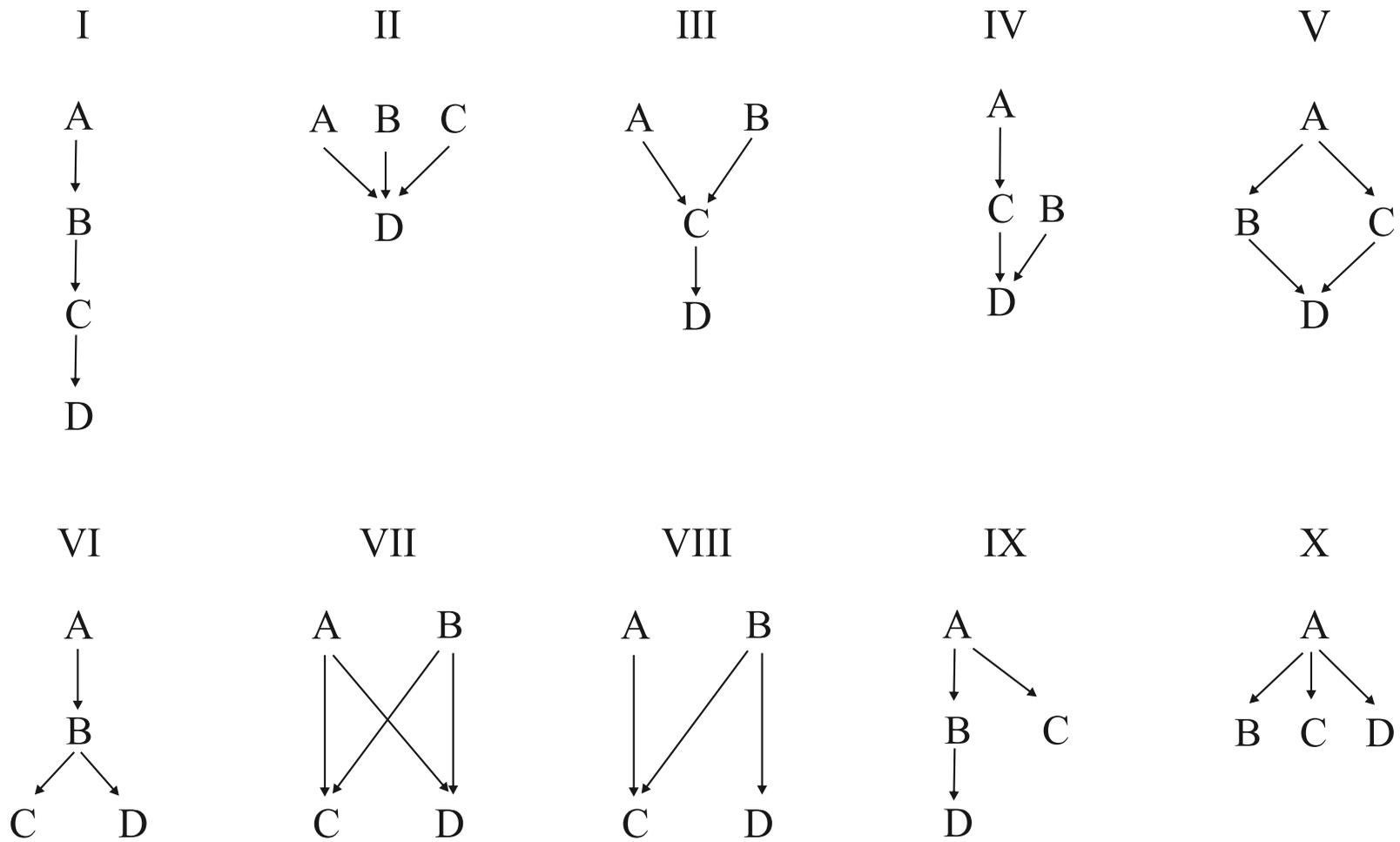
Figure 4. Possible orders with four categories A, B, C, and D. Relations implied by transitivity are not shown.