Journal of Mathematical Psychology 98 (2020) 102378

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

# Hierarchical multinomial modeling to explain individual differences in children's clustering in free recall

Martha Michalkiewicz [a,*,1], Sebastian S. Horn [b,1], Ute J. Bayen [a,1]

[a] Institute for Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany
[b] Department of Psychology, University of Zürich, Binzmühlestr. 14 (Box 11), Zürich, Switzerland

ARTICLE INFO

*Article history:*
Received 14 June 2019
Received in revised form 24 February 2020
Accepted 27 April 2020
Available online xxxx

*Keywords:*
Multinomial processing tree model
Bayesian hierarchical modeling
Individual differences
Cognitive development
Free recall
Category clustering

ABSTRACT

The measurement of individual differences in cognitive processes and the advancement of multinomial processing tree (MPT) models were two of William H. Batchelder's major research interests. Inspired by his work, we investigated developmental differences between 7-year-old children, 10-year-old children, and young adults, in free recall with the pair-clustering model by Batchelder and Riefer (1980, 1986). Specifically, we examined individual differences (in initial levels and in change across multiple study–test trials) in cluster encoding, retrieval, and covariation with three basic cognitive abilities: semantic verbal understanding, short-term memory capacity, information-processing speed. Data from two developmental studies in which 228 participants freely recalled clusterable words in four study–test cycles were used for reanalysis. We combined two model extensions not linked so far (Klauer, 2010; Knapp & Batchelder, 2004). This novel combination of modeling methods made it possible to analyze the relation between individual cognitive abilities and changes in cluster encoding and retrieval across study–test cycles. Inspired by William H. Batchelder, this work illustrates how MPT modeling can contribute to the understanding of cognitive development.



## 1. Introduction

Cognitive psychometrics in the form of multinomial processing tree (MPT) models is one of Bill Batchelder's seminal contributions to mathematical psychology. Through these models, his influence extends far beyond the community of mathematical psychologists into most major subdisciplines of psychology. Having the dissemination of sophisticated mathematical models in substantive areas of psychological research at heart, he established and demonstrated the value of MPT models not only in mainstream cognitive psychology (e.g., Batchelder & Alexander, 2013; Batchelder & Riefer, 1999; Riefer & Batchelder, 1988), but also in the fields of cognitive clinical psychology (Batchelder & Riefer, 2007; Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002), neuropsychology (Batchelder, Chosak-Reiter, Shankle, & Dick, 1997), cognitive aging (Riefer & Batchelder, 1991a), cognitive assessment (e.g., Batchelder, 1998), and social cognition (Batchelder & Batchelder, 2008).

Importantly, he inspired others to do the same leading to innovative developments and applications of MPT models in several areas of memory research (e.g., Bayen, Murnane, & Erdfelder, 1996; Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Erdfelder & Buchner, 1998; Meiser & Bröder, 2002; Riefer & Rouder, 1992; Rummel, Marevic, & Kuhlmann, 2016; Stahl & Klauer, 2008), in research on judgment and decision making (e.g., Heck & Erdfelder, 2017; Hilbig, Erdfelder, & Pohl, 2010; Michalkiewicz & Erdfelder, 2016), cognitive aging (Bayen, Erdfelder, Bearden, & Lozito, 2006; Bayen & Murnane, 1996; Kuhlmann & Touron, 2016; Schnitzspahn, Horn, Bayen, & Kliegel, 2012; Smith, Horn, & Bayen, 2012), child development (Horn, Ruggeri, & Pachur, 2016; Pohl, Bayen, & Martin, 2010; Smith, Bayen, & Martin, 2010), psychopathology (e.g., Groß & Bayen, 2017; Keefe, Arnold, Bayen, & Harvey, 1999; Woodward, Menon, Hu, & Keefe, 2006), psychopharmacology (Walter & Bayen, 2016), social cognition (e.g., Bayen, Nakamura, Dupuis, & Yang, 2000; Klauer & Wegener, 1998; Meissner & Rothermund, 2013), and evolutionary psychology (e.g., Bell & Buchner, 2009; Schaper, Mieth, & Bell, 2019). In areas of inquiry where confounded ad-hoc behavioral measures had previously been used, MPT modeling now allowed researchers to obtain separate and unconfounded measures of cognitive processes of interest. As MPT modeling became popular in cognitive psychology and was introduced in other areas of psychology, Bill Batchelder oversaw the quality of such applications as frequent reviewer for major journals and as associate editor of the Journal of Experimental Psychology: Learning, Memory, and Cognition (2000 to 2003).

* Corresponding author.
*E-mail address:* martha.michalkiewicz@hhu.de (M. Michalkiewicz).

1 We thank with Fabian Gümüsdagli for assistance with data analysis. Sebastian Horn was supported by Swiss National Science Foundation (SNF) grant 100019-185463 for this work.

https://doi.org/10.1016/j.jmp.2020.102378
0022-2496/

The third author of this article first met Bill Batchelder in 1989 when she was a Master's student who, along with her advisor Edgar Erdfelder, sought advice on how to use the pair-clustering model (Batchelder & Riefer, 1980, 1986) in a study on cluster encoding and retrieval in older adults (Erdfelder & Bayen, 1991). This initial encounter with Bill Batchelder and MPT modeling lead to fruitful research activities to the present day. When Bill Batchelder visited us in 2007, he introduced us to novel hierarchical (multi-level) methods in MPT modeling that he published soon thereafter (Smith & Batchelder, 2010). After years of applying MPT models with aggregate data from participant groups that are known to be heterogeneous, including older adults, children, and clinical populations, these new hierarchical methods (Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2015; Smith & Batchelder, 2010) made it possible to estimate parameters for participants and items, thus enabling the investigation of individual differences and correlations with external variables. These novel developments set us on a path to investigate determinants of individual differences in latent cognitive processes (Arnold, Bayen, & Böhm, 2015; Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013; Böhm, Bayen, & Schaper, 2020; Filevich, Horn, & Kühn, 2019; Michalkiewicz, Arden, & Erdfelder, 2018; Schaper, Kuhlmann, & Bayen, 2019) and to use hierarchical MPT models to investigate developmental differences across the lifespan (e.g., Horn, Bayen, & Michalkiewicz, in press; Horn, Pachur, & Mata, 2015; Michalkiewicz, Bayen, & Horn, 2020).

In childhood, there are particularly large individual differences in cognitive abilities. Bill Batchelder's innovative ideas are ideally suited to study interindividual differences in intraindividual change in cognitive development. In this article, we present a novel application of hierarchical MPT modeling to address research questions that have long preoccupied developmental psychologists (e.g., Hasselhorn, 1990; Hünnerkopf, Kron-Sperl, & Schneider, 2009; Moely, Olson, Halwes, & Flavell, 1969): When, how, and why do children learn to cluster related items in free recall? Clustering refers to the formation and encoding of a unit of semantically related items, its maintenance in memory, and its retrieval at recall (Batchelder & Riefer, 1980, 1986). Understanding the development of clustering strategies is important, because clustering is associated with good memory performance (Bjorklund & Jacobs, 1985): Children with higher ability to sort study material according to categories tend to show better memory performance (e.g., Kobasigawa & Middleton, 1972; Moely et al., 1969). To investigate the development of clustering in free recall, we used Batchelder and Riefer's (1980, 1986) MPT model with a multi-trial extension introduced by Knapp and Batchelder (2004) and the hierarchical latent-trait approach introduced by Klauer (2010). This approach allows us to investigate individual differences in clustering and to examine whether basic cognitive abilities may account for these differences. This work is the first to relate external variables to individual learning parameters obtained within a hierarchical MPT modeling framework. We thereby build on two of Bill Batchelder's main research interests, namely, (1) the development and refinement of MPT models and (2) the study of individual differences in cognitive processes.

In what follows, we first sketch the research on differences in clustering in free recall between and within groups of children and young adults. Moreover, we describe how the development of basic cognitive abilities may contribute to individual differences in clustering. We then introduce the pair-clustering MPT model and two modeling extensions and describe their advantages over behavioral measures of clustering. Finally, we show how the modeling approach may help to investigate whether clustering performance can be explained by individual differences in basic cognitive abilities.

## 2. Developmental differences in clustering in episodic free recall

Age-related differences in clustering in free recall from childhood to adulthood are a common finding: younger children typically cluster less than older children, who in turn cluster less than young adults (Bjorklund & Jacobs, 1985; Hünnerkopf et al., 2009; Laurence, 1966; Moely et al., 1969; Schneider & Sodian, 1997). So far, developmental studies have mainly relied on ad-hoc behavioral measures of clustering (e.g., the number of semantically related items recalled adjacently). A common question in such studies has been whether age differences are attributable to encoding processes, retrieval processes, or both (Brainerd, 1985; Brainerd, Howe, Kingma, & Brainerd, 1984; Chechile, Richman, Topinka, & Ehrensbeck, 1981). One important problem with ad-hoc behavioral measures of clustering (Bousfield & Bousfield, 1966; Roenker, Thompson, & Brown, 1971) is that they are a conglomerate of different cognitive processes. For example, the number of related items recalled adjacently (a widely used measure of clustering) confounds encoding and retrieval, because successful adjacent recall of related items involves both of these processes.

The pair-clustering model was proposed by Batchelder and Riefer (1980, 1986) to disentangle encoding and retrieval contributions to clustering in free recall. In two cross-sectional studies with 7-year-old children, 10-year-old children, and young adults, we used this MPT model to investigate age differences in encoding and retrieval processes in free recall of clusterable word lists (Horn et al., in press; Michalkiewicz et al., 2020) and found evidence for age differences in cluster encoding.

Another important question is whether lower levels of clustering in younger children are due to a general inability in strategy use or a lack of experience (Glidden, 1977). Regarding the use of memory strategies, Flavell (1970) assumed a *mediation deficit* in younger children (i.e., a general inability to apply specific strategies) and a *production deficit* in older children (i.e., no spontaneous self-initiated strategy use, but in-principle ability to use memory strategies). In line with this, several studies have shown that younger children do not cluster information in episodic free recall (e.g., Bjorklund, Ornstein, & Haig, 1977), whereas older children do cluster information if additional support (e.g., instruction) is provided (Hasselhorn, 1992; Moely & Jeffrey, 1974). Children may also need more experience than adults to notice the semantic relations between items in a memory task. Therefore, repeated learning opportunities may help children to increasingly apply clustering strategies in memory tasks. Studies using ad-hoc behavioral measures showed differences in clustering between age groups, but findings about changes in clustering with experience are mixed (Cole, Frankel, & Sharp, 1971; Glidden, 1977; Moely & Shapiro, 1971).

The multi-trial extension of the pair-clustering model (see detailed description below) was introduced by Knapp and Batchelder (2004) to quantify change across multiple study–test trials by modeling the dependencies between trials. We used this model version in a previous study to investigate age differences in the initial level of cluster encoding and in the learning rate in cluster encoding across trials (Horn et al., in press).

There is evidence for substantial individual differences in clustering, even in children of similar age (Schneider & Sodian, 1997; Sodian & Schneider, 1999). A likely reason is differences in basic cognitive abilities (e.g., Bjorklund, 1987; Krajewski, Kron, & Schneider, 2004; Richter, 2004). Semantic verbal understanding, short-term memory capacity, and information processing speed provide the basis for successful cluster encoding: To successfully encode two related items as a cluster, a person needs semantic verbal understanding to extract the meaning of the two items and to relate them to each other. Sufficient working-memory capacity allows the person to store the first item of a pair until the second item occurs and a cluster can be formed. Speed of information processing is relevant for the formation and encoding of clusters within the available amount of time. As these basic cognitive abilities develop, clustering is also expected to improve (Bjorklund, 1987; Lange, 1973; Melkman, Tversky, & Baratz, 1981; Richter, 2004; Schleepen & Jonkman, 2012). Until now, the role of basic cognitive abilities in clustering has been mainly investigated with behavioral measures to assess mean-levels of clustering in different age groups. Importantly, however, basic cognitive abilities may also influence the learning progress in clustering that evolves through experience. With repeated learning opportunities, a person may benefit from a higher level of semantic verbal understanding, short-term memory capacity, and information processing speed. A longitudinal study by Richter (2004) provided initial evidence that short-term memory capacity may explain individual differences in the progress of clustering across the elementary-school years.

Based on these considerations, we investigated with a cognitive model whether and how individual differences in cluster encoding and retrieval are related to basic cognitive abilities (semantic verbal understanding, short-term memory capacity, information processing speed). For this goal, we used a multi-trial extension (Knapp & Batchelder, 2004) of the pair-clustering MPT model (Batchelder & Riefer, 1980, 1986), combined with a latent-trait approach (Klauer, 2010).

## 3. Modeling cluster encoding and retrieval

The pair-clustering model (Batchelder & Riefer, 1980, 1986) was the first MPT model developed by Bill Batchelder. Compared to previously used ad-hoc behavioral measures of clustering, a main advantage of this model is that it can disentangle cluster encoding and cluster retrieval involved in recalling pairs of semantically related items. The model is tailored to a free-recall task in which participants are presented with a list of items that includes semantically related pairs. Participants are then asked to recall as many items as possible in any order. As illustrated in Fig. 1, the model accounts for four mutually exclusive response categories (events $E_1$ to $E_4$) that occur in a free-recall task with clusterable item pairs: both items of a pair recalled adjacently ($E_1$); both items of a pair recalled, but non-adjacently ($E_2$); only one of the two items recalled ($E_3$); neither of the items recalled ($E_4$). The frequencies of responses in these categories are modeled through a combination of three parameters: $c$ (the probability of cluster encoding/storage), $r$ (the probability of retrieving a previously encoded cluster), and $u$ (the probability of encoding and retrieving an item as a singleton, if it has not been encoded as part of a cluster). Regarding the model tree, an item pair is encoded as a cluster with probability $c$. With probability $r$, an encoded cluster is retrieved from memory, resulting in both items being recalled adjacently ($E_1$). With complimentary probability $1 - r$, an encoded cluster is not retrieved, resulting in neither of the items being recalled ($E_4$). In case an item pair is not encoded as a cluster (with probability $1 - c$) each of the two items from a pair may be individually encoded and retrieved as a singleton. With probability $u \cdot u$ both items are encoded and retrieved as singletons, resulting in both items being recalled, but non-adjacently ($E_2$). With probability $u \cdot (1 - u)$ only one item of a pair is recalled ($E_3$); with probability $(1 - u) \cdot (1 - u)$ neither of the items is recalled ($E_4$).

The interpretation of the model parameters has been empirically validated (Batchelder & Riefer, 1980; Riefer et al., 2002). The pair-clustering model has helped to better understand group differences in various populations (e.g., Bäuml, 1991; Francis et al., 2018; Golz & Erdfelder, 2004; Riefer & Batchelder, 1991a; Riefer et al., 2002) and also served as a "drosophila" model in methodological advancements of MPT modeling (e.g., Bröder, 2009; Klauer, 2006, 2010; Knapp & Batchelder, 2004; Matzke et al., 2015; Smith & Batchelder, 2010). In its original form, the model was designed to analyze single study–test trials only. Several studies, however, included multiple study–test trials and changes across trials (i.e., learning) were of interest (e.g., Bäuml, 1991; Golz & Erdfelder, 2004; Riefer & Batchelder, 1991a). With the original modeling approach, each of the multiple study–test trials could only be analyzed separately, thereby assuming independence across trials. To characterize change in cognitive processes, it is important to also model dependencies across trials.

## 4. Modeling change

Using a multi-trial modeling approach (Knapp & Batchelder, 2004), we may account for dependencies across trials and estimate the initial level of a given parameter and the within-subject change in this parameter across trials. That is, we can differentiate between the cluster encoding and retrieval probabilities that participants show at the outset of a task and the progress that occurs as a result of learning and experience. The multi-trial approach offers an important modification of the pair-clustering MPT model by specifying order constraints on the parameter values. For the application in a multi-trial learning design with four study–test trials (as in the current paradigm), the model assumes non-decreasing values $p_1 \leq p_2 \leq p_3 \leq p_4$ across trials for each model parameter $p \in \{c, r, u\}$. This change is modeled using the following reparameterization: The model parameter $p \in \{c, r, u\}$ in a given study–test trial $j = 1, \ldots, 4$ is defined as $p_j = 1-(1-p_1)\cdot(1-b_p)^{j-1}$, where $p_1$ is the model parameter on the initial trial and $b_p \in (0, 1)$ is a change parameter.[2] The complementary parameter is defined accordingly as $\left(1 - p_j\right) = (1 - p_1) \cdot (1 - b_p)^{j-1}$. For example, the probability to encode a cluster on the second trial is modeled as $c_2 = c_1 + (1 - c_1) \cdot b_c$, and the complementary probability to not encode a cluster as $(1 - c_2) = (1-c_1)\cdot(1-b_c)$. This results in six model parameters: initial probabilities $c_1, r_1, u_1$, of cluster encoding, retrieval, and recall of single items, respectively, and the rates of change in these parameters across trials: $b_c, b_r, b_u$.

The multi-trial extension of the pair-clustering model has been successfully used to examine the progress of cluster encoding and retrieval across trials at the group level in younger adults, older adults (with and without mild cognitive impairment), and alcoholics (e.g., Bröder, Herwig, Teipel, & Fast, 2008; Knapp & Batchelder, 2004). However, this model version also has limitations. In particular, it is difficult to reliably estimate individual model parameters with relatively few observations per participant (usually between 10 and 20 item pairs in recall studies). Moreover, to examine statistical relations with other variables (e.g., cognitive abilities tests), individual estimates are required. For this purpose, Batchelder and others have proposed hierarchical (multi-level) extensions of MPT models (Klauer, 2010; Matzke et al., 2015; Smith & Batchelder, 2010). In the following, we focus on the latent-trait approach because it is most suitable to address our research aim.

[2] In the current modeling, a constant rate of change across trials was assumed, implying a geometric/exponential learning function (cf. Bush & Mosteller, 1955). This model could account for the data well.

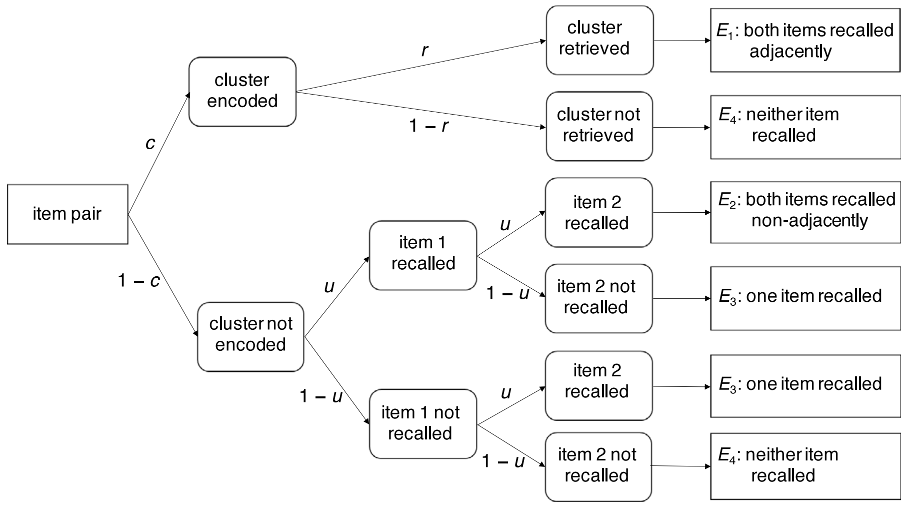


**Fig. 1.** Illustration of the pair-clustering model of free recall of clusterable item pairs (Batchelder & Riefer, 1980, 1986). $c$ = probability of encoding and storing an item pair as a cluster; $r$ = probability of retrieving an encoded and stored cluster; $u$ = probability of recalling (that is, encoding, storing, and retrieving) an item as a singleton. The rectangles in the figure represent observable events (presented items and responses); rectangles with rounded corners represent latent cognitive states.

## 5. Modeling covariates

The hierarchical latent-trait approach (Klauer, 2010) offers two important advantages for the current research aims: First, individual model parameters are estimated more reliably by using the information from the group level, particularly when the numbers of observations per individual are sparse. Second, the approach allows us to model the influence of covariates on the individual parameters by including regression coefficients directly into the estimation of the model parameters. By accounting for the uncertainty in the individual parameter estimates (Katahira, 2016; Matzke et al., 2017), one may assess relations with external variables more accurately than with previous methods (e.g., two-step procedures in which model parameters are estimated in a first step, then followed by separate regression analyses).

Fig. 2 shows the latent-trait multi-trial version of the MPT model for clusterable item pairs (Klauer, 2010; Knapp & Batchelder, 2004). As an illustrative example[3] of how to measure the influence of external variables on model parameters, the three covariates *Similarities* (*SIM*; measuring semantic verbal understanding), *Digit Span* (*DS*; measuring short-term memory capacity), and *Digit–Symbol Coding* (*COD*; measuring information processing speed) are included as predictors of initial cluster encoding (parameter $c_1$). As illustrated in Fig. 2, model parameters are estimated at two levels: The group level comprises the normally distributed group-level means $\mu^p \sim N(0, 1)$ for all parameters $p \in \{c_1, r_1, u_1, b_c, b_r, b_u\}$ and the inverse Wishart distributed variance–covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{-1} \sim \mathrm{W}(\mathbf{I},7)$, estimating the group-level variances $\sigma^p$ and the correlations $\rho^{p\times q}$ between model parameters $p \neq q \in \{c_1, r_1, u_1, b_c, b_r, b_u\}$. The multivariate normally distributed individual deviations from the group mean $\delta_i^p \sim MvN(0, \boldsymbol{\Sigma})$ are derived from $\boldsymbol{\Sigma}$. The uniformly distributed parameters $\xi^p \sim U(0, 100)$ are nuisance parameters included for modeling purposes. At the individual level, each parameter $p_i$ is assessed for each individual $i$ as the individual deviance $\delta_i^p$ (scaled by parameter $\xi^p$) from the group-level mean $\mu^p$: $p_i = \Phi\left(\mu^p + \xi^p \cdot \delta_i^p\right)$, $p \in \{c_1, r_1, u_1, b_c, b_r, b_u\}$. For modeling purposes, the estimation of model parameters is performed in a probit-transformed space (indicated in the equation by the standard normal $\Phi$). For each individual $i$ and study–test trial $j$, category frequencies $\mathbf{C}_{ij} \sim$ Multinomial $(P(\mathbf{C}_{ij}), N_{ij})$ follow a multinomial distribution with probability vector $P(\mathbf{C}_{ij})$ and number of observations $N_{ij}$, defined according to the reparametrized multi-trial version of the pair-clustering model.

Additionally, the influence of the three basic cognitive abilities assessed in our studies can be included into the estimation of the model parameters as regression on the scores of three administered tests: Similarities (*SIM*), Digit Span (*DS*), and Digit–Symbol Coding (*COD*), with the corresponding regression coefficients $\beta_{SIM}$, $\beta_{DS}$, $\beta_{COD}$. As suggested by Heck, Arnold, and Arnold (2018) for small effects, we used a Gamma distribution dgamma(.5, .5*s^2) on the precision parameter of the regression coefficients. This results in the following model equation for each individual $i$ and model parameter $p \in \{c_1, r_1, u_1, b_c, b_r, b_u\}$:

$$p_i = \Phi\left(\mu^p + \xi^p \cdot \delta_i^p + \beta_{SIM}^p \cdot SIM_i + \beta_{DS}^p \cdot DS_i + \beta_{COD}^p \cdot COD_i\right). \quad (1)$$

As an example, Fig. 2 shows how to include the three cognitive variables into the model for initial cluster encoding $c_1$. The corresponding equation reads as follows:

$$c_{1i} = \Phi(\mu^{c_1} + \xi^{c_1} \cdot \delta_i^{c_1} + \beta_{SIM}^{c_1} \cdot SIM_i + \beta_{DS}^{c_1} \cdot DS_i + \beta_{COD}^{c_1} \cdot COD_i). \quad (2)$$

Importantly, only the latent-trait multi-trial pair-clustering model allows us to address our main research aim as it combines the advantages of all approaches explained above: First, it disentangles the encoding and retrieval of clusters. Second, it allows us to assess the initial probabilities of these processes and their change across study–test trials. Third, it enables us to measure cluster encoding and retrieval on an individual level. Fourth, it allows us to assess the role of external variables.

## 6. Reanalysis of data from two developmental studies

We reanalyzed data from two studies in which we investigated category clustering in free recall from episodic memory in school-age children and young adults (Horn et al., in press; Michalkiewicz et al., 2020). In these studies, we also measured participant's basic cognitive abilities (semantic verbal understanding, short-term memory capacity, information-processing

[3] Note that in the reported analyses, the three cognitive variables were included as predictors for *all* parameters in the model, $p \in \{c_1, r_1, u_1, b_c, b_r, b_u\}$. In Fig. 2, this is only illustrated for *one* parameter as an example. Moreover, a factor "Study" and the interactions between this factor and the cognitive variables were included as predictors in the model for all parameters; these additional predictors are also not shown in Fig. 2.

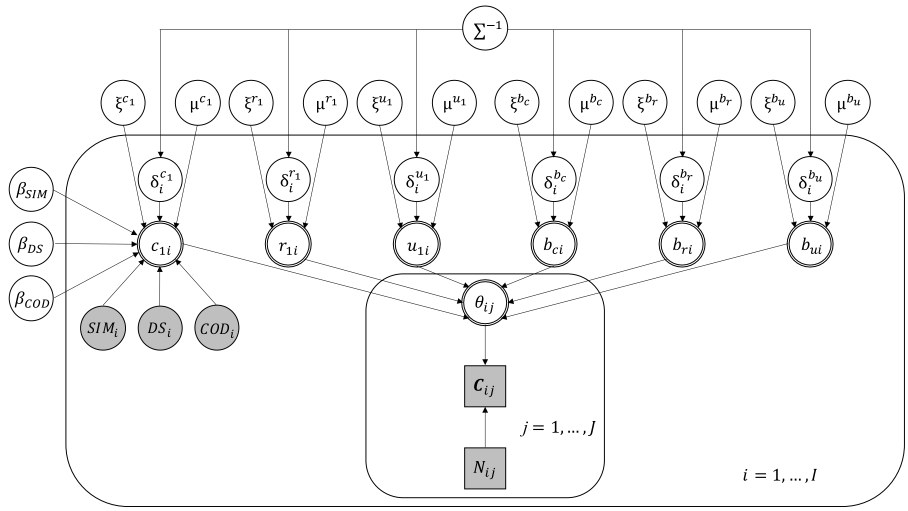


**Fig. 2.** Illustration of the latent-trait multi-trial version of the pair-clustering model for clusterable item pairs (Klauer, 2010; Knapp & Batchelder, 2004). As an example, the external variables Similarities (*SIM*), Digit Span (*DS*), and Digit–Symbol Coding (*COD*) are included as predictors of initial cluster-encoding parameter $c_1$. In line with conventional notation, shaded and unshaded nodes represent observed data and latent model parameters, respectively; square and circular nodes represent discrete and continuous variables, respectively; single-bordered nodes represent model parameters that are estimated from the data, while double-bordered nodes represent model parameters that are computed from the estimated model parameters; the plates represent replications over $I$ individuals and $J = 4$ study–test trials.

**Table 1**
Demographic information and cognitive test scores by age group.

| | 7-year-olds | | 10-year-olds | | Adults | |
|---|---|---|---|---|---|---|
| *N* | 77 | | 68 | | 83 | |
| Gender (female, male) | 44, 33 | | 35, 33 | | 52, 31 | |
| Mean age (years;months) | 6;10 | | 10;3 | | 22;4 | |
| Age range (years;months) | 6;2–7;7 | | 9;3–11;6 | | 18–29 | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Similarities | 10.92 | 3.22 | 11.04 | 2.34 | 9.28 | 2.05 |
| Digit Span | 10.25 | 2.36 | 10.31 | 2.52 | 10.31 | 2.75 |
| Digit–Symbol Coding | 11.19 | 2.53 | 11.38 | 2.42 | 10.82 | 2.79 |

*Notes.* For children, the cognitive tests were taken from the German versions IV and V of the Wechsler Intelligence Scale for Children (Petermann, 2017; Petermann & Petermann, 2007); for adults, the tests were taken from the German version of the Wechsler Adult Intelligence Scale (von Aster, Neubauer, & Horn, 2006). Age-scaled norm scores from Wechsler subtests have $M = 10$ and $SD = 3$; scales range from 1 to 19.

speed) that were not analyzed in the previous papers, but that are of particular interest in the current modeling analysis. We combined the data from both studies to obtain stable estimates in our correlational analyses, which require relatively large sample sizes (see Schönbrodt & Perugini, 2013). In particular, we reanalyzed the data from Horn et al. (in press) and from the control groups of Michalkiewicz et al. (2020) because we used the same research paradigm in both studies and because the sample characteristics, design, and procedures were very similar (see details below). The total sample of $N = 228$ participants included seventy-seven 7-year-olds, sixty-eight 10-year-olds, and eighty-three young adults, all German native speakers. In the few studies using correlational analyses of hierarchical MPT model parameters with external cognitive variables, comparable sample sizes have proven to be adequate to find credible effects (Arnold et al., 2013; Arnold, Bayen, & Smith, 2015; Michalkiewicz et al., 2018; Schaper, Kuhlmann, & Bayen, 2019). Table 1 shows descriptive statistics of demographic and cognitive variables.

In both studies, participants first completed a list-learning task consisting of a study phase, a brief unrelated buffer task (to control for recency effects), and a free-recall phase. In the study phase, we auditorily presented 18 categorically related word pairs (the first four words were excluded from all analyses to control for primacy effects; we used the same two word pairs as a primacy buffer in all study–test trials). In the study by Horn et al. (in press), the two words of each pair were presented separated by one other word (lag 1) or by nine other words (lag 9). In the study by Michalkiewicz et al. (2020), the two words from a pair were presented consecutively (lag 0).[4] To investigate changes in cognitive processes, there were four study–test trials (with a different order of the same study material in each trial). Participants subsequently completed three cognitive tests: Similarities (measuring semantic verbal understanding), Digit Span (measuring short-term memory capacity), and Coding (measuring information processing speed), from the Wechsler intelligence scales for children or adults.

## 7. Results

We first report behavioral measures of recall performance and clustering, followed by cognitive modeling of cluster encoding and retrieval. We analyzed differences both between and within groups. In the modeling, we differentiated between initial cluster encoding and retrieval (study–test trial 1) and the change rate across the four study–test trials to examine whether individual differences were mainly due to clustering that participants showed from the start or that evolved with experience. Finally, and most importantly, we conducted model-based regressions of the MPT parameters on the three cognitive variables (Similarities, Digit Span, and Digit–Symbol Coding) as an attempt to explain individual differences in cluster encoding and retrieval.

Regarding group-level differences in memory performance, in behavioral measures of clustering, and in model-based measures of cluster encoding and retrieval (both initial levels and change rates), the current results obtained with the reanalyses of two

[4] To account for differences between studies, we included a factor "Study" as a covariate and added interaction terms of this factor with the cognitive variables in our MPT model regression analyses. We thank the reviewers for this suggestion. The factor Study did not credibly interact with any of the cognitive-abilities variables (BFs indicated moderate-to-strong evidence for the absence of interactions involving the Study factor), suggesting that the predictive power of the cognitive abilities did not depend on specific task characteristics in the recall paradigm. Tables S1 and S2 in the Supplement provide further details.

combined datasets expectedly mirror the pattern of results reported in the individual studies by Horn et al. (in press) and Michalkiewicz et al. (2020). The important new aspect of the following analyses are the findings regarding individual differences in parameters within age groups and the attempt to explain them through external cognitive variables.

**Age differences in recall performance and clustering.** Table 2 shows descriptive statistics for behavioral measures of recall performance and clustering by age group. There were age differences in the proportion of recalled words and in clustering (assessed as the proportion of word pairs recalled adjacently) both for the initial level on trial 1, $F(2, 225) = 50.58$, $p < .001$, $\eta^2 = .31$, and $F(2, 225) = 29.28$, $p < .001$, $\eta^2 = .21$, and for the change across trials, $F(6, 675) = 91.61$, $p < .001$, $\eta^2 = .45$, and $F(6, 675) = 72.39$, $p < .001$, $\eta^2 = .39$. Adults showed a higher initial level and stronger increases in recalled words and adjacently recalled word pairs than 10-year-olds, who in turn showed a higher initial level and stronger increases than 7-year-olds. The standard deviations within each age group also indicated individual differences. Although the number of word pairs recalled adjacently is a confounded measure of clustering, the behavioral results suggest variability in clustering between and within age groups.

**Model-based analyses.** We investigated the observed differences in clustering in terms of underlying encoding and retrieval processes (separately for the initial level at trial 1 and the change rate across trials) with the latent-trait multi-trial version of the pair-clustering MPT model. We included the influence of the external cognitive variables on each of the model parameters into the model in terms of regressions (as illustrated for parameter $c_1$ in Fig. 2). Additionally, to quantify the evidence for or against an effect, we calculated Bayes factors for the regression coefficients using the Savage–Dickey Method (see Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010), which is defined as the ratio of the density of the posterior and the prior at a slope of zero, and interpreted them according to conventional standards (e.g., Jeffreys, 1961).[5] The reported Bayes factors $BF_{01+}$ represent evidence in favor of $H_0$ (absence of an effect; i.e., a regression coefficient does not differ from zero) over $H_1$ (positive effect; i.e., a regression coefficient is positive). For the analyses of hierarchical MPT models within the Bayesian framework, we used the TreeBUGS package for *R* (Heck et al., 2018). For the models of each age group, we ran six chains with 1,500,000 iterations each, using a thinning rate of 200 and a burn-in period of 500,000. Heterogeneity tests based on the response frequencies in the event categories $E_1$ to $E_4$ in the recall protocols revealed large heterogeneity within each age group and each study–test trial (except for the first trial of the two groups of children): all $\chi^2(228) > 346$, $p < .001$ for 7-year-olds; all $\chi^2(201) > 247$, $p < .02$ for 10-year-olds; all $\chi^2(246) > 405$, $p < .001$ for adults. This large heterogeneity substantiated the necessity to use hierarchical modeling approaches. Chain convergence was satisfactory for all estimated group-level means, standard deviations, and regression coefficients ($\hat{R}$s < 1.05, number of effective samples > 1000; see Gelman, Carlin, Stern, & Rubin, 2004; Kruschke, 2014). Posterior predictive *p* values for the fit indices $T_1$ and $T_2$ (Klauer, 2010) showed an acceptable fit of the model to the mean and covariance structure of the data from the 7-year-olds ($p_{T_1} = .37$; $p_{T_2} = .23$) and the 10-year-olds ($p_{T_1} = .33$; $p_{T_2} = .13$); for adults, the fit to the mean structure was inferior ($p_{T_1} = .004$; $p_{T_2} = .12$). However, plots of the observed data against the posterior-predicted data from the model indicated a satisfactory model fit for all age groups both for the mean and the covariance structure (see Supporting Information, Figs. S1 and S2).

***Age differences in cluster encoding and retrieval.*** Table 2 (bottom half) shows the parameter estimates for the group-level means and standard deviations with Bayesian credibility intervals (BCIs) for the three age groups. Regarding the initial probability ($c_1$) to encode related items as a cluster, young adults did not differ credibly from 7-year-olds, $\Delta_{c_1} = 0.08\,[-0.01, 0.15]$, or 10-year-olds, $\Delta_{c_1} = 0.03\,[-0.06, 0.11]$; the two groups of children also did not differ in initial cluster encoding, $\Delta_{c_1} = -0.05[-0.14, 0.04]$. However, regarding the change rate in clustering across trials ($b_c$), adults showed substantially larger improvement in cluster encoding than 7-year-olds, $\Delta_{b_c} = 0.29$ [0.23, 0.35], and 10-year-olds, $\Delta_{b_c} = 0.25$ [0.18, 0.32]; 7-year-olds did not differ credibly from 10-year-olds, $\Delta_{b_c} = -0.04[-0.08, 0.004]$. Regarding cluster retrieval, there were no credible differences between age groups in initial probabilities ($r_1$): 7-year-olds vs. 10-year-olds: $\Delta_{r_1} = -0.16\,[-0.68, 0.42]$; adults vs. 7-year-olds: $\Delta_{r_1} = 0.41\,[-0.10, 0.80]$; adults vs. 10-year-olds: $\Delta_{r_1} = 0.25\,[-0.12, 0.58]$. Moreover, there were no credible differences in the changes in cluster retrieval across trials ($b_r$) : 7-year-olds vs. 10-year-olds: $\Delta_{b_r} = -0.04\,[-0.72, 0.68]$; adults vs. 7-year-olds: $\Delta_{b_r} = -0.10\,[-0.74, 0.59]$; adults vs. 10-year-olds: $\Delta_{b_r} = -0.14\,[-0.73, 0.52]$. The lack of differences in cluster retrieval must be interpreted with caution due to relatively large uncertainty in retrieval parameter estimates. Taken together, age differences emerged particularly in the rate of change in cluster encoding across trials. We further investigated potential reasons for these differences by regressing parameters on measures of cognitive abilities.

***Role of cognitive abilities in cluster encoding and retrieval.*** Table 3 shows the standardized regression coefficients of the model-based analyses for all parameters. For adults, a credibly positive relation emerged between short-term memory performance (digit span) and the change in encoding related items as clusters (learning across trials; parameter $b_c$). Moreover, higher cognitive speed (as measured with the Digit–Symbol coding test) was associated with better memory for individual items in adults (singleton parameter $u_1$). However, there were no credible relations between initial cluster encoding $c_1$ and measures of cognitive abilities. The corresponding Bayes factors indicated even moderate evidence for no relations (except for 10-year-olds, where BF indicated inconclusive evidence). Moreover, contrary to expectations, there were no credible relations in any age group between performance in the similarities test and initial cluster encoding $c_1$ or the change rate in cluster encoding across trials, $b_c$. Regarding cluster retrieval, there were also no credible relations between the cognitive variables and the model parameters $r_1$ or $b_r$.

## 8. Discussion

We reanalyzed data from two developmental studies in which we investigated categorical clustering in episodic free recall in 7-year-old children, 10-year-old children, and young adults, using a list-learning paradigm with repeated study–test opportunities. One aim of the current study was to examine whether and to what extent basic cognitive abilities (semantic verbal understanding, short-term memory capacity, information processing speed as measured with the Wechsler tests Similarities, Digit Span, and Digit–Symbol Coding, respectively) may explain individual differences in clustering in these three age groups. To obtain measures of the different cognitive components involved in clustering and to assess the specific contribution of the three cognitive variables,

[5] Following conventional interpretation (e.g., Jeffreys, 1961; Lee & Wagenmakers, 2014), in the current analyses, Bayes factors ($BFs_{01}$) between 1/3 and 3 indicate only inconclusive evidence, $BFs_{01}$ between 3 and 10 moderate evidence, and $BFs_{01} > 10$ strong evidence for the absence of an effect. Conversely, $BFs_{01}$ between 1/3 and 1/10 indicate moderate evidence for, and $BFs_{01} < 1/10$ indicate strong evidence for the presence of an effect.

**Table 2**
Recall performance and model parameters by age group.

| | 7-year-olds | 10-year-olds | Adults |
|---|---|---|---|
| Trial | Mean proportion of recalled word (with *SDs*) | | |
| 1 | .14 (.06) | .19 (.07) | .29 (.13) |
| 2 | .20 (.10) | .30 (.09) | .50 (.17) |
| 3 | .23 (.13) | .39 (.12) | .66 (.16) |
| 4 | .25 (.15) | .46 (.14) | .77 (.15) |
| Trial | Mean proportion of adjacently recalled words of a pair (with *SDs*) | | |
| 1 | .02 (.04) | .05 (.06) | .12 (.12) |
| 2 | .07 (.08) | .12 (.11) | .28 (.20) |
| 3 | .07 (.10) | .16 (.12) | .46 (.24) |
| 4 | .10 (.12) | .21 (.14) | .56 (.25) |
| Parameter | Group-level means [with BCIs] | | |
| $c_1$ | .07 [.03, .14] | .12 [.06, .20] | .15 [.10, .21] |
| $r_1$ | .47 [.13, .95] | .64 [.34, .94] | .88 [.66, .99] |
| $u_1$ | .13 [.11, .15] | .15 [.13, .19] | .19 [.16, .22] |
| $b_c$ | .01 [.00, .03] | .05 [.01, .09] | .30 [.24, .35] |
| $b_r$ | .49 [.04, .97] | .53 [.06, .94] | .39 [.04, .87] |
| $b_u$ | .01 [.00, .02] | .05 [.01, .08] | .09 [.04, .14] |
| | Group-level standard deviations [with BCIs] | | |
| $c_1$ | .02 [.00, .07] | .10 [.03, .16] | .12 [.08, .18] |
| $r_1$ | .21 [.00, .45] | .09 [.00, .26] | .07 [.00, .20] |
| $u_1$ | .04 [.02, .06] | .02 [.00, .05] | .07 [.05, .10] |
| $b_c$ | .03 [.00, .07] | .03 [.00, .06] | .16 [.13, .20] |
| $b_r$ | .43 [.26, .48] | .34 [.13, .47] | .17 [.01, .40] |
| $b_u$ | .02 [.01, .06] | .03 [.01, .26] | .08 [.04, .12] |

*Notes.* $c_1$ = probability of encoding a word pair as a cluster in trial 1; $r_1$ = probability of retrieving a stored cluster in trial 1; $u_1$ = probability of encoding and retrieving a word as a singleton; $b_c$ = change in the probability of encoding a word pair as a cluster; $b_r$ = change in the probability of retrieving a word pair as a cluster; $b_u$ = change in the probability of encoding and retrieving a word as a singleton; BCI = Bayesian credibility interval

**Table 3**
Regression Coefficients, Bayesian Credibility Intervals, and Bayes Factors for the Relation Between Cognitive Test Performance and MPT Model Parameters by Age Group.

| | | 7-year-olds | | | 10-year-olds | | | Adults | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Similarities | Digit Span | Coding | Similarities | Digit Span | Coding | Similarities | Digit Span | Coding |
| $c_1$ | $\beta$ | −0.01 | 0.02 | −0.09 | 0.21 | −0.01 | −0.01 | 0.09 | 0.06 | 0.07 |
| | BCI | [−0.27, 0.22] | [−0.25, 0.35] | [−0.34, 0.14] | [−0.04, 0.46] | [−0.32, 0.26] | [−0.27, 0.25] | [−0.10, 0.26] | [−0.12, 0.23] | [−0.16, 0.28] |
| | $BF_{01+}$ | 6.15 | 4.23 | 10.21 | 0.79 | 5.23 | 5.75 | 3.05 | 4.43 | 3.67 |
| $r_1$ | $\beta$ | 0.20 | 0.22 | 0.09 | −0.35 | −0.01 | −0.29 | 0.04 | 0.03 | 0.06 |
| | BCI | [−0.68, 1.01] | [−0.84, 1.29] | [−0.89, 1.03] | [−1.02, 0.32] | [−0.55, 1.10] | [−0.81, 0.31] | [−0.48, 0.59] | [−0.57, 0.60] | [−0.57, 0.68] |
| | $BF_{01+}$ | 1.11 | 0.94 | 1.30 | 4.10 | 1.75 | 5.12 | 2.36 | 2.22 | 1.91 |
| $u_1$ | $\beta$ | −0.004 | −0.05 | 0.01 | 0.06 | −0.02 | −0.02 | 0.01 | 0.05 | **0.14** |
| | BCI | [−0.10, 0.09] | [−0.15, 0.07] | [−0.07, 0.10] | [−0.05, 0.17] | [−0.13, 0.08] | [−0.15, 0.10] | [−0.09, 0.12] | [−0.05, 0.14] | **[0.03, 0.25]** |
| | $BF_{01+}$ | 16.13 | 24.81 | 13.05 | 4.65 | 16.98 | 14.67 | 11.18 | 5.80 | 0.24 |
| $b_c$ | $\beta$ | 0.08 | 0.03 | 0.17 | 0.10 | −0.06 | 0.09 | 0.03 | **0.16** | 0.12 |
| | BCI | [−0.38, 0.56] | [−0.53, 0.61] | [−0.27, 0.62] | [−0.13, 0.37] | [−0.51, 0.37] | [−0.25 0.55] | [−0.12, 0.17] | **[0.01, 0.31]** | [−0.04, 0.29] |
| | $BF_{01+}$ | 2.23 | 2.24 | 1.57 | 2.53 | 3.97 | 2.36 | 6.96 | 0.53 | 1.69 |
| $b_r$ | $\beta$ | 0.29 | −0.12 | 0.09 | −0.39 | 0.09 | 0.05 | 0.17 | −0.24 | −0.09 |
| | BCI | [−1.06, 1.59] | [−1.42, 1.21] | [−1.17, 1.33] | [−1.58, 0.85] | [−0.90, 1.19] | [−1.05, 1.17] | [−0.83, 1.20] | [−1.30, 0.79] | [−1.08, 0.93] |
| | $BF_{01+}$ | 0.72 | 1.20 | 0.98 | 1.82 | 1.19 | 1.19 | 1.05 | 1.88 | 1.61 |
| $b_u$ | $\beta$ | 0.08 | −0.07 | 0.25 | 0.07 | −0.10 | −0.03 | −0.06 | 0.12 | −0.13 |
| | BCI | [−0.33, 0.55] | [−0.56, 0.39] | [−0.12, 0.61] | [−0.28, 0.46] | [−0.34, 0.12] | [−0.34, 0.29] | [−0.42, 0.19] | [−0.10, 0.38] | [−0.41, 0.13] |
| | $BF_{01+}$ | 2.35 | 3.73 | 0.86 | 2.87 | 10.73 | 5.36 | 6.42 | 2.08 | 9.66 |

*Notes.* Coefficients that are credibly different from zero are marked in boldface. $c_1$ = initial cluster encoding; $r_1$ = initial cluster retrieval; $u_1$ = initial encoding and retrieval of a single item; $b_c$ = change in cluster encoding; $b_r$ = change in cluster retrieval; $b_u$ = change in encoding and retrieval of a singleton; BCI = Bayesian credibility interval; $\beta$ = standardized regression coefficients measured within the latent-trait multi-trial version of the pair-clustering model; $BF_{01+}$ = Bayes factor representing evidence in favor of $H_0$ (no effect; regression coefficients being zero) over $H_1$ (positive effect; positive regression coefficients).

we combined two advancements in MPT modeling, the multi-trial approach (Knapp & Batchelder, 2004) and the latent-trait approach (Klauer, 2010), to the pair-clustering model (Batchelder & Riefer, 1980, 1986), resulting in a novel hierarchical implementation. This implementation also included the three cognitive variables and met the current research needs more adequately than previous versions of the pair-clustering model. In particular, the modeling allowed us (1) to disentangle cluster encoding and retrieval, (2) to measure both initial levels and change rates for these processes, (3) to investigate differences between and within age groups, (4) to measure all processes on an individual level, and (5) to relate individual differences in model parameters with external covariates. Thus, the modeling offered important new information by extending the range of research questions that can be addressed.

Regarding developmental differences in cluster encoding (at initial levels and in change rates across study–test trials), the current analyses synthesize and corroborate the findings from two previous developmental studies. In particular, we found evidence for age differences in strategy acquisition (Hünnerkopf et al., 2009; Sodian & Schneider, 1999), that is, in the degree to which clustering-strategy use increased with task experience

(Cole et al., 1971; Nelson, 1969). Regarding the relations between MPT model parameter and measures of cognitive abilities, however, the correlational analyses indicated a pattern that was less clear. Higher cognitive speed was associated with initially better memory for individual items in adults; moreover, higher short-term memory capacity helped people in this age group to progressively encode related items together across study–test trials. However, there was no evidence for similar relations in children. Moreover, contrary to expectations, performance in the similarities test did not account for variability in cluster encoding (parameters $c_1$ and $b_c$). With the help of Bayesian methods, we obtained even some evidence for the absence of relations among the aforementioned variables. What could be potential reasons for this?

First, some developmental studies suggest that measures of basic cognitive abilities may be useful predictors of memory performance—but not necessarily of underlying strategy use in memory tasks (Kron-Sperl, Schneider, & Hasselhorn, 2008). For instance, longitudinal studies have found only low-to-moderate stabilities in strategy use over time, indicating that the development of strategy use may be discontinuous and not follow a monotonically increasing trajectory (Krajewski et al., 2004; Sodian & Schneider, 1999). There is also evidence that children use a mix of old and new strategies concurrently, with these strategies waxing and waning across development in overlapping waves (Siegler, 2016): Many children do not necessarily maintain a memory strategy once they have discovered it, but may drop it and rediscover it at a later stage. This discontinuity does not match the gradual increase across childhood in short-term memory capacity and information processing speed (Dempster, 1981; Kail, 1991) and may reduce the predictive power of cognitive-abilities tests for strategy use.

Second, in line with the notion of a *utilization deficit* (Miller, 1990), strategy use is not necessarily associated with memory performance in younger children (Bjorklund & Coyle, 1995; Miller & Seier, 1994; Schneider & Sodian, 1997; but see Schlagmüller & Schneider, 2002). A common explanation for the utilization deficit is that the application of a strategy and the simultaneous suppression of ineffective strategies exhaust the available cognitive resources, leaving no further resources for active memorization (e.g., Bjorklund & Harnishfeger, 1987; Guttentag, 1984; Miller, 1994). Thus, increases in short-term memory capacity or information processing speed may not necessarily increase children's recall performance when they apply a clustering strategy.

Third, variability in cluster-encoding processes was relatively low in younger children, which makes the detection of relations with other variables difficult. Parameter estimates for cluster encoding (both means and standard deviations) were particularly low in 7-year-olds. In this age group, the probability of encoding related words as a cluster only increased from .07 to .10 across four study–test trials (with minimal within-group variation), suggesting that 7-year-olds did not encode clusters in the first study–test cycle and also did not learn to do so across trials. The task may have been difficult for 7-year-olds, as also indicated by relatively low recall performance (Table 2). It is possible that 7-year-olds may generally not be able to apply clustering strategies in such situations (e.g., Hasselhorn, 1990; Richter, 2004), for instance, due to a limited knowledge base or limited metacognitive skills (Krajewski et al., 2004; Schneider & Sodian, 1997; Sodian & Schneider, 1999). The current findings clearly show that experience and repetition alone are insufficient to induce clustering strategies in younger children. Other types of interventions (e.g., instructional manipulations; Michalkiewicz et al., 2020) or simpler tasks would allow us to examine whether cluster encoding may occur in 7-year-olds, too (for an overview of task simplifications, see Richter, 2004). Despite many advantages of the modeling approach, the current analyses also have limitations. The results regarding cluster retrieval must be interpreted with caution due to large uncertainty in parameter estimates. The problem of estimating the probability of cluster retrieval reliably follows from the low probabilities of cluster encoding (particularly in 7-year-olds) and the structure of the pair-clustering MPT model (Riefer & Batchelder, 1991b): The probability of retrieving a cluster is estimated conditional on the probability that a cluster is successfully encoded (Fig. 1). If the probability of cluster encoding is low (e.g., particularly in younger children), then the number of available observations is too small for reliable estimation of cluster retrieval. Thus, it may not come as a surprise that differences between age groups and/or relations to external cognitive variables could not be reliably discovered for the retrieval parameters. Such difficulties may possibly be overcome by using easier tasks or additional manipulations to increase the level of cluster encoding (Michalkiewicz et al., 2020).

Moreover, our analyses involve a few further challenges. First, we aggregated data from two different studies because correlational analyses require large sample sizes (Schönbrodt & Perugini, 2013). Whereas all procedures in these studies were very similar, the word lists differed in lag between related words in a list. To account for this difference between studies, we included the factor "Study" as a covariate and added interaction terms of this factor with the cognitive variables in the MPT model regression analyses (Tables S1 and S2 in the Supplement provide further details). As can be expected, cluster-encoding parameter values were higher in the study that involved lags of zero exclusively. Importantly, however, the factor Study did not credibly interact with any of the cognitive-abilities variables (BFs indicated moderate-to-strong evidence for the absence of interactions involving the Study factor), suggesting that the predictive power of the cognitive-abilities tests did not depend on the specific lags used in the recall paradigm. Nonetheless, aggregation of data of different lag increased item variability in the current analyses.

Second, some Bayes factors indicated merely inconclusive evidence. In this regard, even larger sample sizes might arguably allow us to arrive at more certain conclusions. Third, further cognitive variables (that we did not consider in the current study) may be more relevant predictors of cluster encoding and retrieval in episodic recall. For example, in the developmental literature, metacognitive knowledge and general knowledge are discussed as further relevant variables that might account for individual differences in clustering (e.g., Krajewski et al., 2004; Schneider & Sodian, 1997; Sodian & Schneider, 1999). Future research should take these considerations into account.

In conclusion, the novel implementation of the pair-clustering model showed that individual differences in younger children's, older children's, and adults' recall of categorically related items are mainly attributable to differences in the learning rate of cluster encoding. This work provides an example of how newest developments in MPT modeling can contribute to the understanding of the development of cognitive processes. The pair-clustering model, most senior among the family of MPT models developed or inspired by Bill Batchelder, is well alive. His legacy lives on in this and other models and in the continuous efforts of those he inspired at every career stage to go beyond the customary, to strive to measure the unobservable, to make assumptions explicit and testable, and to bring new mathematical methods to established fields of inquiry spanning all areas of psychology.

## Appendix A. Online supplemental material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmp.2020.102378. Moreover, scripts and model files can be found at the Open Science Framework at https://tinyurl.com/MPTsRegression.

## References


Arnold, N. R., Bayen, U. J., & Böhm, M. F. (2015). Is prospective memory related to depression and anxiety? A hierarchical MPT modelling approach. *Memory*, *23*, 1215–1228. http://dx.doi.org/10.1080/09658211.2014.969276.

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin and Review*, *20*, 326–333. http://dx.doi.org/10.3758/s13423-012-0342-7.

Arnold, N. R., Bayen, U. J., & Smith, R. E. (2015). Hierarchical multinomial modeling approaches: An application to prospective memory and working memory. *Experimental Psychology*, *62*, 143–152. http://dx.doi.org/10.1027/1618-3169/a000287.

von Aster, M., Neubauer, A., & Horn, R. (2006). *Wechsler Intelligenztest für Erwachsene (WIE) [Wechsler Adult Intelligence Scale, Version III, German Adaptation]*. Frankfurt/Main, Germany: Harcourt Test Services.

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment*, *10*, 331–344. http://dx.doi.org/10.1037/1040-3590.10.4.331.

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

Batchelder, W. H., & Batchelder, E. (2008). Metacognitive guessing strategies in source monitoring. In J. Dunlosky, & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 211–244). New York, NY: Psychology Press, http://dx.doi.org/10.4324/9780203805503.

Batchelder, W. H., Chosak-Reiter, J., Shankle, W. R., & Dick, M. B. (1997). A multinomial modeling analysis of memory deficits in Alzheimer's disease and vascular dementia. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *52*, 206–215. http://dx.doi.org/10.1093/geronb/52B.5.P206.

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review*, *87*, 375–397. http://dx.doi.org/10.1037/0033-295X.87.4.375.

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology*, *39*, 129–149. http://dx.doi.org/10.1111/j.2044-8317.1986.tb00852.x.

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

Batchelder, W. H., & Riefer, D. M. (2007). Using multinomial processing tree models to measure cognitive deficits in clinical populations. In R. W. J. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling of processes and symptoms* (pp. 19–50). Washington, DC: American Psychological Association, http://dx.doi.org/10.1037/11556-001.

Bäuml, K. -H. (1991). Retroaktive Hemmung: Der Einfluß des interpolierten Kategorienmaterials auf die Verfügbarkeit von Information. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *38*, 169–187.

Bayen, U. J., Erdfelder, E., Bearden, J. N., & Lozito, J. P. (2006). The interplay of memory and judgment processes in effects of aging on hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1003–1018. http://dx.doi.org/10.1037/0278-7393.32.5.1003.

Bayen, U. J., & Murnane, K. (1996). Aging and the use of perceptual and temporal information in source memory tasks. *Psychology and Aging*, *11*, 293–303. http://dx.doi.org/10.1037/0882-7974.11.2.293.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197–215. http://dx.doi.org/10.1037/0278-7393.22.1.197.

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. -L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*, 480–500. http://dx.doi.org/10.3758/BF03198562.

Bell, R., & Buchner, A. (2009). Enhanced source memory for names of cheaters. *Evolutionary Psychology*, *7*, 317–330, https://doi.org/10.1177/147470490900700213.

Bjorklund, D. F. (1987). How age changes in knowledge base contribute to the development of children's memory: An interpretive review. *Developmental Review*, *7*, 93–130. http://dx.doi.org/10.1016/0273-2297(87)90007-4.

Bjorklund, D. F., & Coyle, T. R. (1995). Utilization deficiencies in the development of memory strategies. In F. R. Weinert, & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 161–180). London, England: Psychology Press.

Bjorklund, D. F., & Harnishfeger, K. K. (1987). Developmental differences in the mental effort requirements for the use of an organizational strategy in free recall. *Journal of Experimental Child Psychology*, *44*, 109–125. http://dx.doi.org/10.1016/0022-0965(87)90025-7.

Bjorklund, D. F., & Jacobs, J. W. (1985). Associative and categorical processes in children's memory: The role of automaticity in the development of organization in free recall. *Journal of Experimental Child Psychology*, *39*, 599–617. http://dx.doi.org/10.1016/0022-0965(85)90059-1.

Bjorklund, D. F., Ornstein, P. A., & Haig, J. R. (1977). Developmental differences in organization and recall: Training in the use of organizational techniques. *Developmental Psychology*, *13*, 175–183. http://dx.doi.org/10.1037/0012-1649.13.3.175.

Böhm, M. F., Bayen, U. J., & Schaper, M. L. (2020). Are subjective sleepiness and sleep quality related to prospective memory? *Cognitive Research: Principles and Implications*, *5*(5), https://doi.org/10.1186/s41235-019-0199-7.

Bousfield, A. K., & Bousfield, W. A. (1966). Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, *19*, 935–942. http://dx.doi.org/10.2466/pr0.1966.19.3.935.

Brainerd, C. J. (1985). Model-based approaches to storage and retrieval development. In C. J. Brainerd, & M. Pressley (Eds.), *Basic processes in memory development* (pp. 143–207). New York, NY: Springer, http://dx.doi.org/10.1007/978-1-4613-9541-6_4.

Brainerd, C. J., Howe, M. L., Kingma, J., & Brainerd, S. H. (1984). On the measurement of storage and retrieval contributions to memory development. *Journal of Experimental Child Psychology*, *37*, 478–499. http://dx.doi.org/10.1016/0022-0965(84)90072-9.

Bröder, A. (2009). Semantically clustered words are stored with integrated context: Validating a measurement model for source memory, storage, and retrieval in free recall. *Zeitschrift für Psychologie*, *217*, 136–148. http://dx.doi.org/10.1027/0044-3409.217.3.136.

Bröder, A., Herwig, A., Teipel, S., & Fast, K. (2008). Different storage and retrieval deficits in normal aging and mild cognitive impairment: A multinomial modeling analysis. *Psychology and Aging*, *23*, 353–365. http://dx.doi.org/10.1037/0882-7974.23.2.353.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, *124*, 137–160. http://dx.doi.org/10.1037/0096-3445.124.2.137.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. Oxford, England: John Wiley & Sons, Inc, http://dx.doi.org/10.1037/14496-000.

Chechile, R. A., Richman, C. L., Topinka, C., & Ehrensbeck, K. (1981). A developmental study of the storage and retrieval of information. *Child Development*, *52*, 251–259. http://dx.doi.org/10.2307/1129238.

Cole, M., Frankel, F., & Sharp, D. (1971). Development of free recall learning in children. *Developmental Psychology*, *4*, 109–123. http://dx.doi.org/10.1037/h0030435.

Dempster, F. N. (1981). Memory span: sources of individual and developmental differences. *Psychological Bulletin*, *89*, 63–100. http://dx.doi.org/10.1037/0033-2909.89.1.63.

Erdfelder, E., & Bayen, U. J. (1991). Episodisches Gedächtnis im Alter: Methodologische und empirische Argumente für einen Zugang über mathematische Modelle [Episodic memory in old age: Methodological and empirical arguments for a mathematical modeling approach]. In D. Frey (Ed.), *Bericht über den 37. Kongress der Deutschen Gesellschaft für Psychologie in Kiel 1990* (pp. 172–180). Göttingen, Germany: Hogrefe.

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 387–414. http://dx.doi.org/10.1037/0278-7393.24.2.387.

Filevich, E., Horn, S. S., & Kühn, S. (2019). Within-person adaptivity in frugal judgments from memory. *Psychological Research*, *83*, 613–630, https://doi.org/10.1007/s00426-017-0962-7.

Flavell, J. H. (1970). Developmental studies of mediated memory. *Advances in Child Development and Behavior*, *5*, 181–211. http://dx.doi.org/10.1016/S0065-2407(08)60467-X.

Francis, W. S., Taylor, R. S., Gutiérrez, M., Liaño, M. K., Manzanera, D. G., & Penalver, R. M. (2018). The effects of bilingual language proficiency on recall accuracy and semantic clustering in free recall output: evidence for shared semantic associations across languages. *Memory*, *26*, 1364–1378. http://dx.doi.org/10.1080/09658211.2018.1476551.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Texts in statistical science series*, *Bayesian data analyses*. Boca Raton, FL: Chapman & Hall/CRC, http://dx.doi.org/10.1002/sim.1856.

Glidden, L. M. (1977). Developmental effects in free recall learning. *Child Development*, *48*, 9–12. http://dx.doi.org/10.2307/1128874.

Golz, D., & Erdfelder, E. (2004). Effekte von L-Dopa auf die Speicherung und den Abruf verbaler Informationen bei Schlaganfallpatienten [Effects of L-Dopa on storage and retrieval of verbal information in stroke patients]. *Zeitschrift für Neuropsychologie*, *15*, 275–286, http://dx.doi.org/10.1024/1016-264X.15.4.275.

Groß, J., & Bayen, U. J. (2017). Effects of dysphoria and induced negative mood on the processes underlying hindsight bias. *Cognition and Emotion*, *31*, 1715–1724. http://dx.doi.org/10.1080/02699931.2016.1249461.

Guttentag, R. E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology*, *37*, 92–106. http://dx.doi.org/10.1016/0022-0965(84)90060-2.

Hasselhorn, M. (1990). Kategoriales Organisieren als Gedächtnisstrategie: Allgemeine und differentielle Entwicklungsperspektiven im Grundschulalter [Categorical organizing as memory strategy: general and differential developmental perspectives in elementary-school age]. In M. Knopf, & W. Schneider (Eds.), *Entwicklung. Allgemeine Verläufe - Individuelle Unterschiede - Pädagogische Konsequenzen* (pp. 117–143). Göttingen, Germany: Hogrefe.

Hasselhorn, M. (1992). Entwicklung kategorialen Organisierens als Gedächtnisstrategie: Zur Rolle des Aufgabenkontextes und der Interitem-Assoziativität [Development Of categorical organizing as memory strategy: the role of task context and interitem associativity]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *24*, 317–334, Retrieved from https://www.pedocs.de/volltexte/2013/1704/pdf/Hasselhorn_1992_Entwicklung_kategorialen_Denkens.pdf.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*, 264–284. http://dx.doi.org/10.3758/s13428-017-0869-7.

Heck, D. W., & Erdfelder, E. (2017). Linking process and measurement models of recognition-based decisions. *Psychological Review*, *124*, 442–471. http://dx.doi.org/10.1037/rev0000063.

Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010). One-reason decision making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 123–134. http://dx.doi.org/10.1037/a0017518.

Horn, S. S., Bayen, U. J., & Michalkiewicz, M. (2020). The development of clustering in episodic memory: A cognitive-modeling approach. *Child Development*.

Horn, S. S., Pachur, T., & Mata, R. (2015). How does aging affect recognition-based inference? A hierarchical Bayesian modeling approach. *Acta Psychologica*, *154*, 77–85. http://dx.doi.org/10.1016/j.actpsy.2014.11.001.

Horn, S. S., Ruggeri, A., & Pachur, T. (2016). The development of adaptive decision making: Recognition-based inference in children and adolescents. *Developmental Psychology*, *52*, 1470–1485. http://dx.doi.org/10.1037/dev0000181.

Hünnerkopf, M., Kron-Sperl, V., & Schneider, W. (2009). Die Entwicklung des strategischen Gedächtnisses im Laufe der Grundschulzeit: Zusammenfassende Ergebnisse der Würzburger Längsschnittstudie [Development of strategic memory over the course of elementary school: summarized results from the Würzburg longitudinal study]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *41*, 1–11. http://dx.doi.org/10.1026/0049-8637.41.1.1.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, *109*, 490–501. http://dx.doi.org/10.1037/0033-2909.109.3.490.

Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, *73*, 37–58. http://dx.doi.org/10.1016/j.jmp.2016.03.007.

Keefe, R. S. E., Arnold, M. C., Bayen, U. J., & Harvey, P. D. (1999). Source monitoring deficits in patients with schizophrenia; a multinomial modelling analysis. *Psychological Medicine*, *29*, 903–914. http://dx.doi.org/10.1017/S0033291799008673.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*, 7–31. http://dx.doi.org/10.1007/S11336-004-1188-3.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*, 70–98. http://dx.doi.org/10.1007/S11336-009-9141-0.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology*, *75*, 1155–1178. http://dx.doi.org/10.1037/0022-3514.75.5.1155.

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. http://dx.doi.org/10.1016/j.jmp.2004.03.002.

Kobasigawa, A., & Middleton, D. B. (1972). Free recall of categorized items by children at three grade levels. *Child Development*, *43*, 1067–1072. http://dx.doi.org/10.2307/1127659.

Krajewski, K., Kron, V., & Schneider, W. (2004). Entwicklungsveränderungen des strategischen Gedächtnisses beim Übergang vom Kindergarten in die Grundschule. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *36*, 47–58. http://dx.doi.org/10.1026/0049-8637.36.1.47.

Kron-Sperl, V., Schneider, W., & Hasselhorn, M. (2008). The development and effectiveness of memory strategies in kindergarten and elementary school: Findings from the Würzburg and Göttingen longitudinal memory studies. *Cognitive Development*, *23*, 79–104. http://dx.doi.org/10.1016/j.cogdev.2007.08.011.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Cambridge, MA: Academic Press, http://dx.doi.org/10.1016/C2012-0-00477-2.

Kuhlmann, B. G., & Touron, D. R. (2016). Aging and memory improvement through semantic clustering: The role of list-presentation format. *Psychology and Aging*, *31*, 771–785. http://dx.doi.org/10.1037/pag0000117.

Lange, G. (1973). The development of conceptual and rote recall skills among school age children. *Journal of Experimental Child Psychology*, *15*, 394–406. http://dx.doi.org/10.1016/0022-0965(73)90090-8.

Laurence, M. W. (1966). Age differences in performance and subjective organization in the free-recall learning of pictorial material. *Canadian Journal of Psychology*, *20*, 388–399. http://dx.doi.org/10.1037/h0082958.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*, 205–235. http://dx.doi.org/10.1007/s11336-013-9374-9.

Matzke, D., Ly, A., Selker, R., Weeda, W. D., Scheibehenne, B., Lee, M. D., & Wagenmakers, E.-J. (2017). Bayesian inference for correlations in the presence of measurement error and estimation uncertainty. *Collabra: Psychology*, *3*(25), http://dx.doi.org/10.1525/collabra.78.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 116–137. http://dx.doi.org/10.1037/0278-7393.28.1.116.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the implicit association test: The real model for the IAT. *Journal of Personality and Social Psychology*, *104*, 45–69. http://dx.doi.org/10.1037/a0030734.

Melkman, R., Tversky, B., & Baratz, D. (1981). Developmental trends in the use of perceptual and conceptual attributes in grouping, clustering, and retrieval. *Journal of Experimental Child Psychology*, *31*, 470–486. http://dx.doi.org/10.1016/0022-0965(81)90031-X.

Michalkiewicz, M., Arden, K., & Erdfelder, E. (2018). Do smarter people employ better decision strategies? The influence of intelligence on adaptive use of the recognition heuristic. *Journal of Behavioral Decision Making*, *31*, 3–11. http://dx.doi.org/10.1002/bdm.2040.

Michalkiewicz, M., Bayen, U. J., & Horn, S. S. (2020). *Modeling the influence of instructions on children's clustering in free recall*. Manuscript in preparation.

Michalkiewicz, M., & Erdfelder, E. (2016). Individual differences in use of the recognition heuristic are stable across time, choice objects, domains, and presentation formats. *Memory & Cognition*, *44*, 454–468. http://dx.doi.org/10.3758/s13421-015-0567-6.

Miller, P. H. (1990). The development of strategies of selective attention. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development* (pp. 157–184). Hillsdale, NJ: Lawrence Erlbaum Associates, http://dx.doi.org/10.4324/9780203771648.

Miller, P. H. (1994). Individual differences in children's strategic behaviors: Utilization deficiencies. *Learning and Individual Differences*, *6*, 285–307. http://dx.doi.org/10.1016/1041-6080(94)90019-1.

Miller, P. H., & Seier, W. L. (1994). Strategy utilization deficiencies in children: When, where, and why. In H. W. Reese (Ed.), *Advances in child development and behavior*: *Vol. 25*, (pp. 107–156). San Diego, CA: Academic Press, http://dx.doi.org/10.1016/S0065-2407(08)60051-8.

Moely, B. E., & Jeffrey, W. E. (1974). The effect of organization training on children's free recall of category items. *Child Development*, *45*, 135–143. http://dx.doi.org/10.2307/1127759.

Moely, B. E., Olson, F. A., Halwes, T. G., & Flavell, J. H. (1969). Production deficiency in young children's clustered recall. *Developmental Psychology*, *1*, 26–34. http://dx.doi.org/10.1037/h0026804.

Moely, B. E., & Shapiro, S. I. (1971). Free recall and clustering at four age levels: Effects of learning to learn and presentation method. *Developmental Psychology*, *4*(490). http://dx.doi.org/10.1037/h0030960.

Nelson, K. J. (1969). The organization of free recall by young children. *Journal of Experimental Child Psychology*, *8*, 284–295. http://dx.doi.org/10.1016/0022-0965(69)90103-9.

Petermann, F. (2017). *Wechsler Intelligenztest für Kinder V (WISC-V) [Wechsler intelligence scale for children, version V, German adaptation]*. Frankfurt am Main, Germany: Pearson Assessment.

Petermann, F., & Petermann, U. (2007). *Hamburg-Wechsler-Intelligenztest für Kinder IV (HAWIK-IV) [Wechsler intelligence scale for children, version IV, German adaptation]*. Bern, Switzerland: Huber.

Pohl, R. F., Bayen, U. J., & Martin, C. (2010). A multiprocess account of hindsight bias in children. *Developmental Psychology*, *46*, 1268–1282. http://dx.doi.org/10.1037/a0020209.

Richter, M. (2004). *Nutzung und Effektivität der kategorialen Organisationsstrategie im Grundschulalter (Doctoral dissertation) [Use and effectiveness of categorial organization in elementary school age]*. Retrieved from Deutsche Nationalbibliothek (https://dnb.de).

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339. http://dx.doi.org/10.1037/0033-295x.95.3.318.

Riefer, D. M., & Batchelder, W. H. (1991a). Age differences in storage and retrieval: A multinomial modeling analysis. *Bulletin of the Psychonomic Society*, *29*, 415–418. http://dx.doi.org/10.3758/BF03333957.

Riefer, D. M., & Batchelder, W. H. (1991b). Statistical inference for multinomial processing tree models. In J. -P. Doignon, & J. -C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 313–335). New York, NY: Springer, http://dx.doi.org/10.1007/978-1-4613-9728-1_18.

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*, 184–201. http://dx.doi.org/10.1037/1040-3590.14.2.184.

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*, 601–611. http://dx.doi.org/10.3758/BF03202710.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*, 45–48. http://dx.doi.org/10.1037/h0031355.

Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1526–1543. http://dx.doi.org/10.1037/xlm0000266.

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 470–496. http://dx.doi.org/10.1037/xlm0000602.

Schaper, M. L., Mieth, L., & Bell, R. (2019). Adaptive memory: Source memory is positively associated with adaptive social decision making. *Cognition*, *186*, 7–14. http://dx.doi.org/10.1016/j.cognition.2019.01.014.

Schlagmüller, M., & Schneider, W. (2002). The development of organizational strategies in children: Evidence from a microgenetic longitudinal study. *Journal of Experimental Child Psychology*, *81*, 298–319. http://dx.doi.org/10.1006/jecp.2002.2655.

Schleepen, T. M. J., & Jonkman, L. M. (2012). Children's use of semantic organizational strategies is mediated by working memory capacity. *Cognitive Development*, *27*, 255–269. http://dx.doi.org/10.1016/j.cogdev.2012.03.003.

Schneider, W., & Sodian, B. (1997). Memory strategy development: Lessons from longitudinal research. *Developmental Review*, *17*, 442–461. http://dx.doi.org/10.1006/drev.1997.0441.

Schnitzspahn, K. M., Horn, S. S., Bayen, U. J., & Kliegel, M. (2012). Age effects in emotional prospective memory: cue valence differentially affects the prospective and retrospective component. *Psychology and Aging*, *27*, 498–509. http://dx.doi.org/10.1037/a0025021.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*, 609–612. http://dx.doi.org/10.1016/j.jrp.2013.05.009.

Siegler, R. S. (2016). Continuity and change in the field of cognitive development and in the perspectives of one cognitive developmentalist. *Child Development Perspectives*, *10*, 128–133. http://dx.doi.org/10.1111/cdep.12173.

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

Smith, R. E., Bayen, U. J., & Martin, C. (2010). The cognitive processes underlying event-based prospective memory in school age children and young adults: A formal model-based study. *Developmental Psychology*, *46*, 230–244. http://dx.doi.org/10.1037/a0017100.

Smith, R. E., Horn, S. S., & Bayen, U. J. (2012). Prospective memory in young and older adults: The effects of ongoing-task load. *Aging, Neuropsychology, and Cognition*, *19*, 495–514. http://dx.doi.org/10.1080/13825585.2011.633161.

Sodian, B., & Schneider, W. (1999). Memory strategy development—gradual increase, sudden insight, or roller coaster? In F. E. Weinert & W. Schneider (Ed.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 61–77). Cambridge, UK: Cambridge University Press.

Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 570–586. http://dx.doi.org/10.1037/0278-7393.34.3.570.

Wagenmakers, E. -J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. *Cognitive Psychology*, *60*, 158–189. http://dx.doi.org/10.1016/j.cogpsych.2009.12.001.

Walter, N. T., & Bayen, U. J. (2016). Selective effects of acute alcohol intake on the prospective and retrospective components of a prospective-memory task with emotional targets. *Psychopharmacology*, *233*, 325–339. http://dx.doi.org/10.1007/s00213-015-4110-z.

Woodward, T. S., Menon, M., Hu, X., & Keefe, R. S. (2006). Optimization of a multinomial model for investigating hallucinations and delusions with source monitoring. *Schizophrenia Research*, *85*, 106–112. http://dx.doi.org/10.1016/j.schres.2006.03.008.