Memory Representations, Tree Structures, and Parameter Polysemy: Comment on Cooper, Greve, and Henson (2017)

David Kellen

Syracuse University

Henrik Singmann

University of Zurich

Author Note

David Kellen, Department of Psychology. Henrik Singmann, Department of Psychology. Correspondence should be sent to David Kellen (davekellen@gmail.com).
We thank Elisa Cooper for making her data available. David Kellen and Henrik Singmann received support from the Swiss National Science Foundation Grant 100014_165591.

Abstract

Cooper, Greve, and Henson (2017) discussed the use of different approaches for measuring item and source memory, and how choices among these can affect the comparison between different groups (e.g., younger versus older adults). The authors argue that the tree structure adopted in the specification of item- and source-memory retrieval in multinomial processing tree models implies a theoretical commitment to the way memories are represented. According to the authors, this commitment can affect the conclusions that are taken from the results and produce different model fits in experimental designs involving confidence-rating judgments. Reported model fits suggest that an alternative tree structure provides a superior account of the data. The present comment argues that the particular tree structure used does not enforce any commitment to a particular structure, as long as the trees are well defined in the sense that parameters are monosemic across the entire structure of the model. The different fit results reported by CGH are due to their reliance on tree structures that do not ensure that all parameters are monosemic.

Introduction

In the study of human memory, researchers often call upon the distinction between individuals' ability to remember the context and characteristics of previously encountered stimuli — a faculty known as *source memory* — and their ability to merely remember that stimuli were encountered before — *item memory* (e.g., Brainerd, Reyna, Holliday, & Nakamura, 2013; Johnson, Hashtroudi, & Lindsay, 1993; Hautus, Macmillan, & Rotello, 2008; Klauer & Kellen, 2010; Onyper, Zhang, & Howard, 2010). Several studies have employed this distinction and assessed how item and source memory are affected by different experimental variables (e.g., stimulus similarity; Bayen, Murnane, & Erdfelder, 1996) and individual covariates (e.g., age; Ferguson, Hashtroudi, & Johnson, 1992).

Cooper, Greve, and Henson (2017; henceforth CGH) discussed different ways in which item and source memory can be measured. Their general take-home message was that no measure is theory-free, and that it is the researcher's responsibility to evaluate the extent to which their results hinge on the adopted measures' theoretical assumptions (for similar points, see Bröder & Meiser, 2007; Rotello, Heit, & Dube, 2014). CGH's discussion focuses on measurements obtained with multinomial processing tree (MPT) models (for reviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009), a class of measurement models that has long been associated to the study of item and source memory (e.g., Batchelder & Riefer, 1990).

One of the claims made by CGH is that the tree structure of the MPT model used by researchers implies a commitment to how memories are represented. The model is typically specified in a way such that source memory is measured via the estimated probability that source memory is available *conditional* on the presence of item memory. CGH refer to this model as the *item-source model*. According to CGH, this model implicitly assumes a single memory representation that at a minimum supports item memory, but with sufficient quality is able to support both item and source memory (see the left panel of Figure 1). CGH argue that by relying on different tree structure, one can establish an alternative model, which they refer to as the *source-item model*,

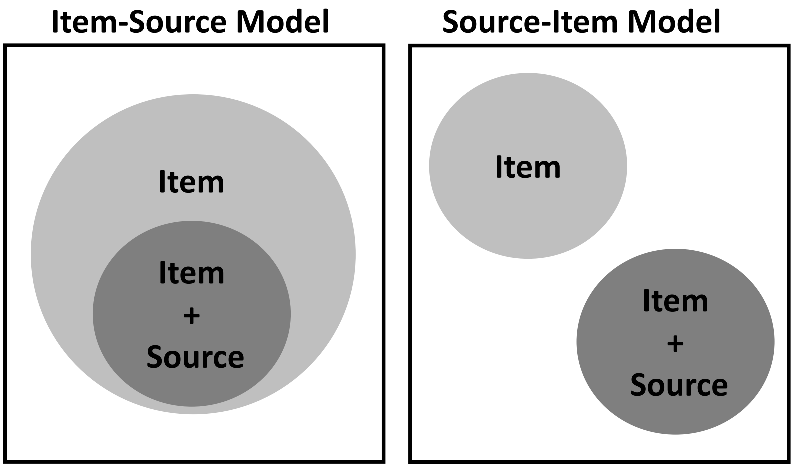

Figure 1. Different memory representations as considered by Cooper, Greve, and Henson (2017).

that is in line with the notion that there are two distinct representations, one supporting item memory, and another supporting both item and source memory (see the right panel of Figure 1). This alternative model, which is reported to provide a superior fit of the data, is argued to be more in line with neuroimaging data supporting the notion of two distinct representations.

The claim that different tree structures embody different memory representations is somewhat surprising given that the opposite scenario has been well documented in the literature (e.g., Buchner & Erdfelder, 1996; Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Hu, 2001). CGH admit that different memory representations cannot be distinguished by tree structures when individuals provide binary responses, but argue that a distinction is possible when confidence judgments are also included. CGH's argument is supported by different goodness-of-fit results for the item-source model compared to the source-item model for data with confidence judgments (their Experiment 2). However, CGH fail to inform the reader of the reasons why confidence judgments allow for such a discrimination.

As will be shown below, the MPT model used to describe item and source memory makes no assumptions whatsoever on the existence of one or two types of memory representations, even when applied to an extended design using a confidence-rating scale. The differences in goodness of fit reported by CGH result from the particular tree structures they adopted, which not only differ drastically from the ones used in previous work the (e.g., Klauer & Kellen, 2010), but also have problematic features such as parameter polysemy (i.e., parameters have a different meaning in different parts of the model). The remainder of this comment is organized as follows: We will first discuss the high-threshold model of source memory and its extension to a confidence-rating paradigm. This discussion will be made along the lines of previous work with this model (e.g., Klauer & Kellen, 2010), which CGH do not appear to have followed. We will then fit the data from CGH's Experiment 2 and show that changes in the tree structure do no affect the model's ability to characterize data. Finally, we will discuss the problems associated with the tree structures used by CGH. Importantly, these problems reinforce the message that CGH were trying to convey: one should always keep in mind what parameters mean in the context of the model that they are part of.

High-Threshold Models of Item and Source Memory

In a typical source-memory task, participants study stimuli from two sources, A and B, which differ on one or more dimensions (e.g., color, position, gender/race of the speaker, etc.). During the test phase, participants are presented with an intermixed list of old and new items. For each test item, participants are requested to judge whether a stimulus was studied in one of the sources (responses "A" and "B") or not at all (response "New").

The high-threshold model originally proposed by Batchelder and Riefer (1990) and later extended by Bayen et al. (1996) attempts to provide a principled decomposition of the observed responses in terms of a mixture of memory- and guessing-based processes. This decomposition is achieved by assuming that the tested stimuli are mapped onto a finite set of discrete mental states M. In this particular two-source task, the following mental states can be distinguished:

- M_1: For an A item, it is remembered that the item is old and from Source A.

- M_2: For a B item, it is remembered that the item is old and from Source B.
- M_3: For an old item, it is remembered that the item is old, but memory for the source is absent.
- M_4: For a new item, it is detected that it is new.
- M_5: For an item presented at test, it is not remembered that the item is old, and it is not detected that the item is new.

Note that state M_3 corresponds to a situation of partial information given that the individual can determine the tested stimulus as old but not its source. In the case of state M_5, there is no information available whatsoever (i.e., there is complete information loss; see Kellen & Klauer, 2015).

The probability of each mental state being entered, given a particular stimulus (A, B, or new), is determined by a *stimulus-state mapping function* π. For example, for A stimuli, we can specify $\pi_{M_1|A}$, $\pi_{M_3|A}$, and $\pi_{M_5|A}$, as the probabilities of a Source A stimulus entering mental states M_1, M_3, and M_5, respectively (with $\pi_{M_1|A} + \pi_{M_3|A} + \pi_{M_5|A} = 1$). As shown by Hu and Batchelder (1994), this mapping function can be always be represented by a binary tree structure without any loss of generality. The binary tree structure shown in Figure 2 is one such representation in which $\pi_{M_1|A} = D_A \times d_A$, $\pi_{M_3|A} = D_A \times (1 - d_A)$, and $\pi_{M_5|A} = 1 - D_A$. Parameter D_A corresponds to the probability that item memory is retrieved, whereas d_A corresponds to the probability of source memory being retrieved, conditional on the fact that item memory was retrieved. Analogous parameters D_B and d_B can be established for M_2 and M_3, as well as a parameter D_N that captures the probability of a new item being actively rejected as new (i.e., entering M_4).

At this point, it is easy to see how using different tree structures does not change the model in any meaningful way (e.g., change any representational assumptions). For example, one could follow CGH and build a so-called source-item model that instead assumes that $\pi_{M_1|A} = d'_A$, $\pi_{M_3|A} = (1 - d'_A) \times D'_A$, and $\pi_{M_5|A} = (1 - d'_A) \times (1 - D'_A)$, where d'_A is the probability that item and source memory are retrieved, and D'_A is the

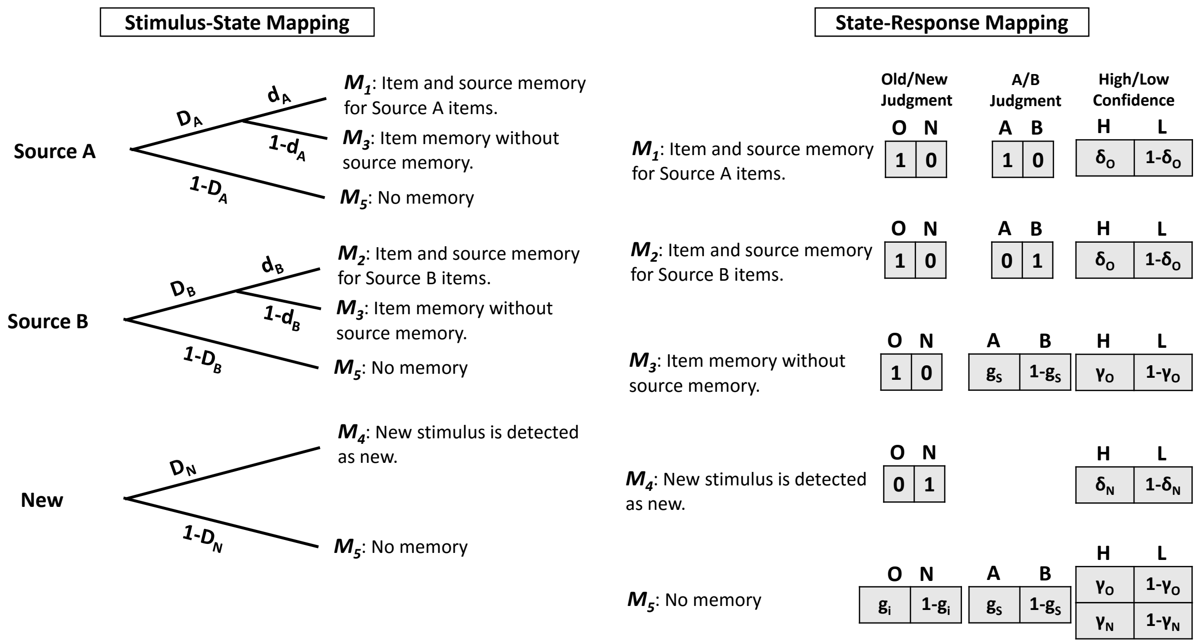

Figure 2. Stimulus-state and state-response mapping functions of the high-threshold model. D_A = Probability of detecting a Source A item as old. d_A = Probability of detecting a Source A item as coming from Source A, conditional on remembering it as old. D_B = Probability of detecting a Source B item as old. d_B = Probability of detecting a Source B item as coming from Source B, conditional on remembering it as old. D_N = Probability of detecting a new item as new. g_I = Probability of guessing a stimulus as old, conditional on the stimulus not being remembered. g_S = Probability of guessing a stimulus as coming from Source A, conditional on the stimulus being deemed old. δ_O = Probability of producing a high-confidence response, given that the stimulus was detected as coming from either Source A or B. δ_N = Probability of producing a high-confidence response, given that the stimulus was detected as new. γ_O = Probability of producing a high-confidence response, given that the stimulus was guessed to be old. γ_N = Probability of producing a high-confidence response, given that the stimulus was guessed to be new.

probability that item memory is retrieved, conditional on the failure to retrieve both item and source memory. This parametrization simply represents a different way to express the exact same thing, namely the probabilities of each mental state M being reached.[1]

When discussing different tree structures, CGH argue that the different tree structures yield parameters with different meanings. Although this is entirely true, it is important to keep in mind that ultimately there is no information gain nor loss when using a specific parametrization versus another. One can easily transform the parameter values from one parametrization into the other parameterization (e.g., $d'_A = D_A \times d_A$).

[1]Ultimately, one could simply set aside these parametrizations and estimate the stimulus-state mapping parameters π directly (e.g., Hu & Batchelder, 1994).

The only difference is that the parameter estimates one obtains might be more or less amenable to the comparisons researchers want to do. For example, if one wants to study the relationship between some covariates and the probability of both item and source memory being retrieved, then it might be more *useful* to rely on a parametrization that sets $\pi_{M_1|A} = d'_A$ rather than $\pi_{M_1|A} = D_A \times d_A$ as one has to deal with a single parameter rather than the product of two.

State-Response Mapping

Establishing the different states and how they can be reached is not sufficient, as one still needs to specify the observed responses that result from these states — i.e., a *state-response mapping function*. In cases where both item and source memory is retrieved, or a new item is actively rejected, it is assumed that individuals respond correctly with probability 1. When partial information or no information are available, the individual is assumed to rely on guessing processes, guessing whether the stimulus is "old" or "new" with probabilities g_i and $1 - g_i$, respectively. Moreover, when a stimulus is deemed to be old (either due to item memory or via guessing) but no source memory is available, the individual guesses "A" or "B" with probabilities g_s and $1 - g_s$, respectively.

If the experimental design also includes confidence-rating judgments, the state-response mapping function needs to be extended accordingly. Specifically, we need to establish the probabilities with which the different confidence levels of a given response can be reached. For example, when individuals are requested to report low or high confidence for their response, we need to consider the probabilities with which an individual who reaches mental state M_1 responds "A" with low and high confidence. Figure 2 illustrates this extended response mapping using different parameters for the different mental states M. It is sometimes assumed that certain states should be assumed to always generate high confidence responses (e.g., Yonelinas, 1999). However, these additional constraints are ultimately inadequate as they impose something that ultimately is not part of the core principles of the model (see Klauer & Kellen, 2010;

Malmberg, 2002). Also, the state-mapping functions should be flexible enough to capture different response styles participants might have (e.g., use of extreme ratings) rather than conflate specific response categories with postulated processes (Bockenholt, 2016; Klauer & Kellen, 2010). More reasonable constraints on state-response mapping functions (e.g., inequality constraints) that still allow for responses to be distributed across the confidence scale have been shown to provide good accounts of the data (e.g., Klauer & Kellen, 2010; Klauer & Kellen, 2015), and focused experimental tests have shown that these state-response mappings can be selectively influenced (e.g., Bröder, Kellen, Schütz, & Rohrmeier, 2013; Kellen, Singmann, Vogt, & Klauer, 2015).[2]

One important aspect that will be critical below is the notion of *parameter monosemy*. Each parameter in the model has exactly the same meaning across the whole model. For example, in all trees. parameter g_i corresponds to the probability of guessing old, conditional on the absence of any memory information. As will be discussed below, the differences reported by CGH result from having parameters that are *polysemic*; the meaning of the parameter changes across trees within the same model.

Fitting CGH's Experiment 2 Data

We will now fit the confidence-rating data from younger and older adults from CGH's Experiment 2. We will fit the high-threshold model illustrated in Figure 2, which establishes parameters for source retrieval conditional on the retrieval of item memory — an item-source model according to CGH's terminology. In line with CGH's analysis we will restrict the detection parameters for the old stimuli, D_{BT} and d_{BT}, to be equal across the two sources. We will also fit a variant which specifies a parameter for the joint retrieval of item and source memory (see the description above): $\pi_{M_1|BT} = d'_{BT}$, $\pi_{M_3|BT} = (1 - d'_{BT}) \times D'_{BT}$, and $\pi_{M_5|BT} = (1 - d'_{BT}) \times (1 - D'_{BT})$. In line with CGH's terminology we will refer to this second model as the source-item model.

[2]Selective influence tests not only show that one can manipulate the state-response mapping parameters (e.g., change how individuals use the scale) without affecting the stimulus-state parameters, but also that we can manipulate memory performance without affecting the state-response mapping parameters (Province & Rouder, 2012).

Table 1

Results of Fitting the Traditional Source-Memory Model for Confidence Rating to CGH's Experiment 2

Younger Adults											
Item-Source Model											
Goodness of Fit				Mean Parameters							
Summed G^2	*df*	*p*-value	Significant misfits	D_{BT}	d_{BT}	D_N	g_I	g_S	δ_{ON}	γ_O	γ_N
91.45	84	.27	3/12 (= 25%)	.83	.80	.49	.36	.37	.81	.33	.63
Source-Item Model											
Goodness of Fit				Mean Parameters							
Summed G^2	*df*	*p*-value	Significant misfits	D'_{BT}	d'_{BT}	D_N	g_I	g_S	δ_{ON}	γ_O	γ_N
91.45	84	.27	3/12 (= 25%)	.57	.65	.49	.36	.37	.81	.33	.63
Older Adults											
Item-Source Model											
Goodness of Fit				Mean Parameters							
Summed G^2	*df*	*p*-value	Significant misfits	D_{BT}	d_{BT}	D_N	g_I	g_S	δ_{ON}	γ_O	γ_N
94.26	84	.21	3/12 (= 25%)	.88	.54	.56	.32	.47	.89	.56	.67
Source-Item Model											
Goodness of Fit				Mean Parameters							
Summed G^2	*df*	*p*-value	Significant misfits	D'_{BT}	d'_{BT}	D_N	g_I	g_S	δ_{ON}	γ_O	γ_N
94.26	84	.21	3/12 (= 25%)	.76	.49	.56	.32	.47	.89	.56	.67

Note. Subscript BT denotes the detection parameters set equal across Bottom and Top stimuli. Column "Significant misfits" reports the number of individuals fits (each with df=7) for which the computed G^2 statistic yielded a p-value below .05. δ_{ON} is the probability of a high confidence response given that the item was detected as old and its source remembered or the item was detected as new. γ_O is the probability of a high confidence response given that the item was guessed to be old. γ_N is the probability of a high confidence new response given that the item was guessed to be new. In order to sidestep unstable parameter estimates due to zero cells, the mean estimates reported were obtained with data to which $\frac{1}{6}$ was added to each cell (this amounts to adding a single additional trial to each stimulus type).

The task used in CGH's Experiment 2 involved individuals judging whether a stimulus was previously presented in the bottom (B) or top (T) part of the screen. In order to fit this model we also imposed a restriction on the state-response mapping functions. Specifically, we set $\delta_O = \delta_N$; due to this equality we will refer to parameter δ_{ON} below. For each participant, CGH's experimental design yields $3 \times 5 = 15$ degrees of freedom, which outnumber the 8 parameters estimated by the model. Note that the model fitted by CGH to this data also employed 8 parameters. Fits were conducted with `R` package `MPTinR` (Singmann & Kellen, 2013) using the maximum-likelihood method.

Fit results and parameter estimates are reported in Table 1. The so-called item-source and source-item models produced exactly the same fits and all but the detection parameters for the old stimuli (i.e., D_{BT} and d_{BT} versus D'_{BT} and d'_{BT}) were identical, as expected. Overall, model fits indicate that were no gross mischaracterizations of the data (three statistically-significant misfits per age group). Mean parameter estimates for the item-source model suggest that younger and older adults have comparable probabilities of retrieving item memory (Wilcoxon $W = 71$, $p = .98$), but the probabilities of further retrieving source memory are smaller for the latter group ($W = 108$, $p = .04$). The parameters of the source-item model suggest that younger adults have a lower probability of retrieving both item and source memory, but conditional on the failure of this joint retrieval, older adults seem to be more likely to retrieve item memory alone. However, none of these differences seem to be systematic (largest $W = 94$, smallest $p = .24$). The fact that the statistical results from these tests differ is not surprising, after all both models assume different parameters and therefore partition the uncertainty in the data differently. Whereas the item-source memory estimates source-memory retrieval conditional on the retrieval of item memory, the source-item memory estimates the joint retrieval of item and source memory. When attempting to evaluate differences in source memory alone across groups, the parametrization of the item-source model appears to be more suitable, as it does not conflate source-memory retrieval with the retrieval of item memory.

The state-response mapping parameters, which are identical for both models, show that individuals tend to map memory-based responses onto higher-confidence judgments than guessing-based responses. This pattern, which is expected to hold under minimal theoretical assumptions (individuals should be more sure of their responses when they are based on memory information rather than guessing), replicates previously-published studies (Bröder et al., 2013; Kellen et all., 2015; Klauer & Kellen, 2010).

What is Behind the Differences Reported by CGH?

Given the equivalence between both models, we now need to understand what drove the differences in fit reported by CGH. Because their respective tree structures

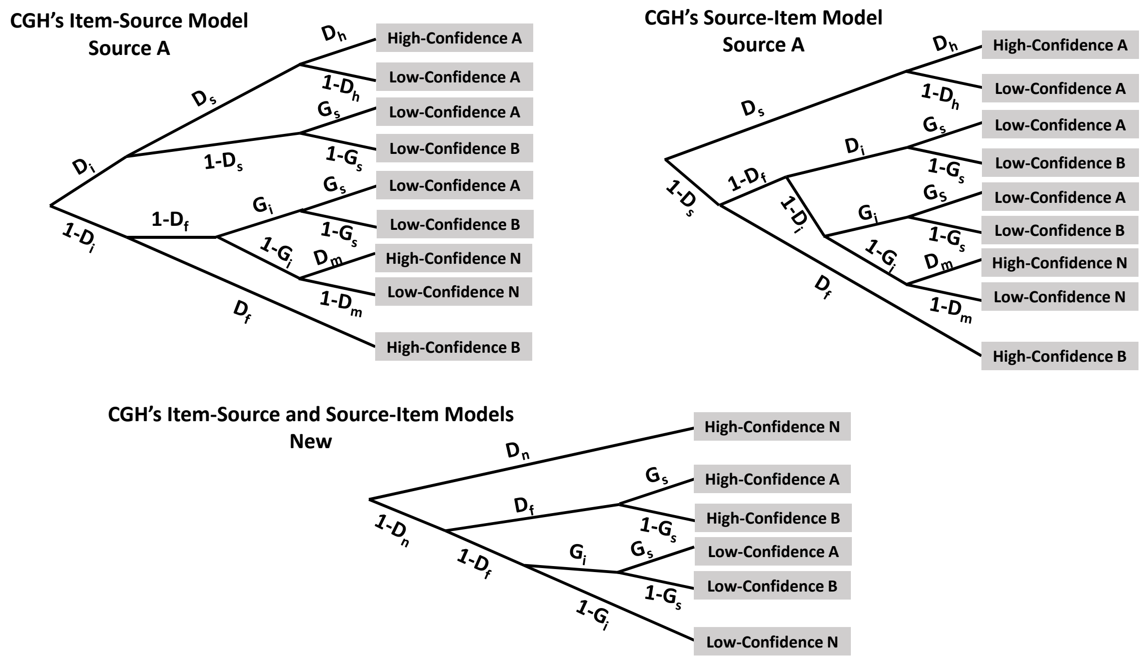

Figure 3. Illustration of Cooper, Greve, and Henson's (2017) item-source and source-item models. CGH's parameter definition: D_s = probability of remembering source (assumed to be equal for both sources). D_i = Remembering item as old (assumed to be equal for both sources). G_s = Guessing an item's source as A. G_i = Guessing an item as old. D_n = Concluding that an unstudied item is new. D_h = Veridical recollection. D_f = False recollection. D_m = Missed encoding

.

deviate from traditional specifications in non-trivial ways, it is informative to go through it in detail. Figure 3 provides a detailed description of their item-source and source-item models, including their parameter nomenclature.

Figure 3 describes two models reflecting the item-source and source-item models as specified by CGH (we only discuss the trees for Source A and new stimuli as these two suffice). One important aspect of the two models is that they include the possibility of "false recollection" measured via parameter D_f, which is invariably mapped onto incorrect source judgments with maximum confidence. In fact, all of CGH's models assume that such high-confidence errors can only result from false recollection. This assumption is much stricter to what has been previously done in the literature:

Although many before have associated recollection processes to maximum-confidence judgments, they nevertheless allowed for other processes (e.g., guessing) to be mapped onto maximum-confidence judgments as well (e.g., Yonelinas, 1999). CGH do not make such concessions, but also do not provide any justification for their stronger assumptions.[3]

The different ways in which D_f is specified in the two models is what is behind the differences in fit. In the Source A tree of the item-source model, D_f corresponds to the probability that there is false recollection conditional on the absence of any kind of veridical memory (the branch is $(1 - D_i) \times D_f$). The exact same notion applies the New stimulus tree (e.g., $(1 - D_n) \times D_f \times G_s$). However, this is not the case in CGH 's source-item model, where the meaning of D_f depends on the tree: In the Source A tree, D_f represents the probability of false recollection conditional on the absence of a joint retrieval item and source memory (the branch is $(1 - D_s) \times D_f$), such that the retrieval of item memory (captured by parameter D_i) is yet to be determined. In the New stimulus tree, D_f again represents the probability of false recollection conditional on the absence of any kind of veridical retrieval. The polysemy of D_f in CGH's source-item model is instrumental in its apparent superiority over the item-source model. When using CGH's formulation across age groups, the performance of source-item model (summed $G^2 = 242.48$, $df = 168$, $p < .001$) is superior to the item-source model performance of source-item model (summed $G^2 = 304.02$, $df = 168$, $p < .001$), a difference that is statistically significant (Wilcoxon test, $W = 28$, $p = .004$).

If one adjusts CGH's source-item model such that D_f always represents the probability of false recollection in the absence of any veridical memory (e.g., we would have $(1 - D_s) \times (1 - D_i) \times D_f$ instead of $(1 - D_s) \times D_f$), then the source-item model again always yields fits that are exactly equal to the item-source model. Note that the superior fit of CGH's "uncorrected" source-item model should not interpreted as

[3]Another questionable aspect of D_f is that for old items it always leads to incorrect source judgments (capturing the notion that the source is incorrectly recollected), but in the case of new items, individuals still have to engage in a guessing process to determine the source of the stimulus. The meaning of D_f is again not well defined. Note that previous models including false-recollection parameters (e.g., Brainerd et al., 2013) do not have this issue as they do not assume that such a process can take place with new items.

evidence that this model provides a better characterization of the data. Ultimately the validity of this model is questionable as it relies on stimulus-state and state-response mapping functions that are difficult to justify. As forcefully discussed by Roberts and Pashler (2000), the goodness of fit of a model (even after correcting for flexibility) is only one of the criteria based on which it should be judged. Other criteria such as theoretical soundness and parameter validity should also be factored in.[4]

Discussion

CGH argued that a source-memory MPT model with an alternative tree structure, where source is not dependent on the retrieval of item memory, yields not only different but better results than the traditional version of the model. As we showed here, the two tree structures are mere reparametrizations of the same model, and do not commit to any particular representation of memory. It turns out that the reported differences in fit reported by CGH are due to the peculiarities of the tree structures they considered, namely the introduction of a polysemic parameter in one of the models. This issue bring us back to a key message of CGH's paper, which is that the meaning of each parameter (i.e., measure) should be carefully considered and clarified when constructing and implementing any model.

The equivalence between different tree structures (when all parameters are monosemic) raises the question of whether one can compare different conceptualizations of how memory processes are arranged (e.g., do they follow some specific serial order?). Fortunately we can, but the conditions that permit it are not trivial and specific experimental designs often need to be employed (e.g., Schweickert & Han, 2016; Schweickert & Chen, 2008; Schweickert & Xi, 2011). Another promising approach involves the incorporation of response times into the MPT model by estimating the duration times associated with the different tree branches (Hu, 2001). Similar opportunities seem to arise when combining behavioral and neural data, as shown

[4]Even if one would focus on fit performance alone – something we do not advocate – there would still be no support for CRH's source-item model. We compared it with the traditional specification of the high-threshold model used in our reanalysis. According to the Fisher Information Approximation, a model selection statistic that takes the functional flexibility of models into account, CRH's source-item model was found to perform worse in 92% of the individual datasets.

recently by Anderson, Zhang, Borst, and Walsh (2016).

References

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of Sternberg's method. *Psychological Review*, *123*, 481–509. doi:10.1037/rev0000030

Batchelder, W. H. & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548–564. doi:10.1037//0033-295X.97.4.548

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

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. doi:10.1037//0278-7393.22.1.197

Bockenholt, U. (2016). Measuring response styles in likert items. *Psychological Methods*. doi:10.1037/met0000106

Brainerd, C. J., Reyna, V. F., Holliday, R. E., & Nakamura, K. (2012). Overdistribution in source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 413–439. doi:10.1037/a0025645

Bröder, A., Kellen, D., Schütz, J., & Rohrmeier, C. (2013). Validating a two-high threshold model for confidence rating data in recognition memory. *Memory*, *8*, 916–944. doi:10.1080/09658211.2013.767348

Bröder, A. & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*, 52–60. doi:10.1027/0044-3409.215.1.52

Buchner, A. & Erdfelder, E. (1996). On assumptions of, relations between, and evaluations of some process dissociation measurement models. *Consciousness and Cognition*, *5*, 581–594. doi:http://dx.doi.org/10.1006/ccog.1996.0034

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. doi:10.1037/0096-3445.124.2.137

Cooper, E., Greve, A., & Henson, R. N. (2017). Assumptions behind scoring source versus item memory: effects of age, hippocampal lesions and mild memory problems. *Cortex*. doi:http://dx.doi.org/10.1016/j.cortex.2017.01.001

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, *217*, 108–144. doi:10.1027/0044-3409.217.3.108

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology and aging*, *7*(3), 443–452. doi:10.1037/0882-7974.7.3.443

Hautus, M., Macmillan, N. A., & Rotello, C. M. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *15*, 889–905. doi:10.3758/PBR.15.5.889

Hu, X. (2001). Extending general processing tree models to analyze reaction time experiments. *Journal of Mathematical Psychology*, *45*, 603–634. doi:http://dx.doi.org/10.1006/jmps.2000.1340

Hu, X. & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the em algorithm. *Psychometrika*, *59*(1), 21–47. doi:10.1007/BF02294263

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. doi:10.1037//0033-2909.114.1.3

Kellen, D. & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review*, *122*, 542–557. doi:10.1037/a0039251

Kellen, D., Singmann, H., Vogt, J., & Klauer, K. C. (2015). Further evidence for discrete-state mediation in recognition memory. *Experimental Psychology*, *62*, 40–53. doi:10.1027/1618-3169/a000272

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

Klauer, K. C. & Kellen, D. (2015). The flexibility of models of recognition memory: the case of confidence ratings. *Journal of Mathematical Psychology*, *67*, 8–25. doi:http://dx.doi.org/10.1016/j.jmp.2015.05.002

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 380–387. doi:10.1037//0278-7393.28.2.380

Onyper, S., Zhang, Y., & Howard, M. W. (2010). Some-or-none recollection: Evidence for item and source memory. *Journal of Experimental Psychology: General*, *139*, 341–362. doi:10.1037/a0018926

Province, J. M. & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences USA*, *109*, 14357–14362. doi:10.1073/pnas.1103880109

Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367. doi:10.1037//0033-295X.107.2.358

Rotello, C. M., Heit, E., & Dubé, C. (2015). When more data steer us wrong: Replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review*, *22*, 944–954.

Schweickert, R. & Chen, S. (2008). Tree inference with factors selectively influencing processes in a processing tree. *Journal of Mathematical Psychology*, *52*, 158–183. doi:http://dx.doi.org/10.1016/j.jmp.2008.01.004

Schweickert, R. & Han, H. J. (2016). Two topics in tree inference: locating a phonological network effect in immediate recall and arborescence partitive set form. *Journal of Mathematical Psychology*, *71*, 7–21. doi:http://dx.doi.org/10.1016/j.jmp.2016.02.001

Schweickert, R. & Xi, Z. (2011). Multiplicatively interacting factors selectively influencing parameters in multiple response class processing and rate trees.

Journal of Mathematical Psychology, *55*, 348–364. doi:http://dx.doi.org/10.1016/j.jmp.2011.06.002

Singmann, H. & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models with R. *Behavior Research Methods*, *45*, 560–575. doi:10.3758/s1342801202590

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1415–1434. doi:10.1037/0278-7393.25.6.1415