Registered Report

Testing Hypotheses About Binding in Context Memory With a Hierarchical Multinomial Modeling Approach

A Preregistered Study

Nina R. Arnold, Daniel W. Heck, Arndt Bröder, Thorsten Meiser, and C. Dennis Boywitt

Department of Psychology, School of Social Sciences, University of Mannheim, Germany

Abstract: In experiments on multidimensional source memory, a stochastic dependency of source memory for different facets of an episode has been repeatedly demonstrated. This may suggest an integrated representation leading to mutual cuing in context retrieval. However, experiments involving a manipulated reinstatement of one source feature have often failed to affect retrieval of the other feature, suggesting unbound features or rather item-feature binding. The stochastic dependency found in former studies might be a spurious correlation due to aggregation across participants varying in memory strength. We test this artifact explanation by applying a hierarchical multinomial model. Observing stochastic dependency when accounting for interindividual differences would rule out the artifact explanation. A second goal is to elucidate the nature of feature binding: Contrasting encoding conditions with integrated feature judgments versus separate feature judgments are expected to induce different levels of stochastic dependency despite comparable overall source memory if integrated representations include feature-feature binding. The experiment replicated the finding of stochastic dependency and, thus, ruled out an artifact interpretation. However, we did not find different levels of stochastic dependency between conditions. Therefore, the current findings do not reveal decisive evidence to distinguish between the feature-feature binding and the item-context binding account.

Keywords: context memory, context binding, multinomial modeling, individual differences, hierarchical modeling, Bayesian inference

PREREGISTERED

OPEN DATA

Source memory is referred to as remembering the origin of information (e.g., Johnson, Hashtroudi, & Lindsay, 1993). Successful source memory is especially important to evaluate the quality or trustworthiness of information. For example, you might want to remember who gave you an important health advice – your doctor or your hairdresser. The context of the learning episode, however, is typically multi-faceted, i.e., it is characterized by a multitude of different features. Concerning the example, you might not remember the person giving you the health advice, but you remember where you heard it – in the doctor's office or in the beauty salon. Situations with more than one context dimension are valuable for investigating the *binding* of context features in source memory. It has been argued that features bound into one episode in memory will show mutual cuing and thus stochastic dependence in retrieval success. Stochastic dependence in context retrieval is present when retrieval of one context feature (e.g., spatial position) is more likely if the other context feature (e.g., gender) is successfully retrieved (e.g., Meiser & Bröder, 2002). Stochastic dependence has been shown across several experiments and manipulations (for an overview see, e.g., Hicks & Starns, 2015), but there are different alternative explanations of this phenomenon.

The aim of the current work is to test different accounts in a pre-registered experiment, using hierarchical modeling as a means for safeguarding against a spurious correlation. First, we will briefly describe the competing accounts before we describe the method and analysis of the experiment.

Explanations of Stochastic Dependence in Context Retrieval

The *mutual cuing hypothesis* (Meiser & Bröder, 2002) argues that successfully retrieving one context feature serves as cue for the other context feature if both have been encoded into the same memory trace. However, experimentally reinstating one context dimension during test (as a means to boost retrieval of this feature) did not lead to better remembering of the other feature in several experiments (Hicks & Starns, 2015, 2016; Starns & Hicks, 2005, 2008; Vogt & Bröder, 2007). Additionally, stochastic dependency between two features has also been shown when both features were tested in separate test phases, speaking against mutual cuing during retrieval (Starns & Hicks, 2005).

In contrast, the *binding-variability hypothesis* (Starns & Hicks, 2005; 2008) states that stochastic dependence is a matter of covariation on factors that influence the binding of the respective context features to the item due to participant- or item-related sources of variation (e.g., attention). According to this reasoning, stochastic dependence in multidimensional source retrieval arises because stronger (or weaker) item-context binding for one context dimension coincides with stronger (or weaker) item-context binding on other context dimensions across the items of the learning list. Observed stochastic dependence is thus assumed to reflect simultaneous variations in source memory accuracy on the item level without direct binding among context features (Hicks & Starns, 2015).

Stochastic dependence was observed specifically in the state of "remembering" an item in the remember/know paradigm (Tulving, 1985) but not in the state of "knowing" an item (Meiser & Bröder, 2002). In the remember/know paradigm, participants are asked whether they consciously remember having seen the item (*R* responses) or simply have a feeling of familiarity ("knowing the item", *K* responses). When aggregating over R and K responses separately, R responses usually show higher overall source memory than K responses, which might indicate that differences in stochastic dependence actually result from general source memory strength rather than subjective R and K memory states. However, Meiser, Sattler, and Weisser (2008) showed that stochastic dependence persisted only in R responses but not in K responses even when memory performance for R and K responses was equated experimentally.

Boywitt and Meiser (2012a) further tested implications of the *context-context binding* assumption of stochastic dependence and suggested that the encoding process is crucial for context-context binding because the context features must be integrated into one memory trace to be bound together. Accordingly, stochastic dependence on the basis of context-context binding should only be observed if two context features co-occur during the encoding episode, whereas variations in item-context binding can lead to stochastic dependence even if item-context relations are learned in separate learning phases (Hicks & Starns, 2015). The authors therefore varied whether the context features were presented simultaneously or separately during the encoding phase and found no stochastic dependence when context features were presented separately. In another study with incidental source memory instructions, Boywitt and Meiser (2012b) found stochastic dependence only for R responses in an intrinsic condition (in which features were part of the to-be-studied item) as opposed to an extrinsic condition (in which features were part of the surroundings of the item). With explicit source memory instructions, R responses showed binding regardless of whether the context was intrinsic or extrinsic. The authors concluded that context features must be in the focus of attention during encoding to be encoded in a bound fashion.

To summarize, stochastic dependency in multidimensional source retrieval has been shown various times and has been interpreted as a result of mutual cuing among context features in memory by some authors (e.g., Boywitt & Meiser, 2012a, 2012b; Meiser & Bröder, 2002). Starns and Hicks (2005, 2008), in contrast, suggested that the effect may reflect a spurious correlation due to variations in item-context binding between items. Thus, the purpose of this study is to test whether binding of context features can be demonstrated by dissociating stochastic dependence from source memory strength and whether binding can be manipulated independently of source memory.

For separating source memory and binding, we employ the multinomial model of memory for multidimensional source information (Meiser & Bröder, 2002; Meiser, 2014). Previously, the model has only been applied to data aggregated across items and participants. Another purpose of this study therefore is to address whether stochastic dependence can be attributed to a spurious correlation due to data aggregation and is driven by individual differences in overall context memory. According to this alternative explanation, some participants have better memory for both context features than other participants, in which case aggregating over participants (and items) leads to spurious stochastic dependency. This point was initially addressed by Meiser et al. (2008) by analyzing subgroups of participants with similar levels of overall source memory. Here, we apply a hierarchical version of the multinomial model that we describe in more detail in the next section. This methodological improvement resolves the possible artifact of aggregation by disentangling the stochastic dependency in retrieving multiple context attributes for a given item from interindividual differences in source memory performance for various context features.

The Hierarchical Multinomial Model of Memory for Multidimensional Source Information

Based on the two-high-threshold model of source monitoring (2HTSM, Bayen, Murnane, & Erdfelder, 1996), Meiser and Bröder (2002) developed a multinomial processing tree (MPT) model for two crossed dimensions of source information that provides independent measures of item and source memory as well as guessing. Meiser (2014) introduced a re-parameterized version of the model yielding separate parameters specifying stochastic dependence and stochastically independent retrieval, respectively. The model was experimentally validated by testing the selective influence of source similarity on distinct dimensions (Meiser & Bröder, 2002). Several studies (e.g., Boywitt & Meiser 2012a, 2012b; Meiser, 2014) found stochastic dependence only for items that were consciously remembered by the participants (R responses) but not for items that were just claimed familiar (K responses). This suggests that the model is sensitive to changes in context binding.

The re-parameterized model variant by Meiser (2014) is displayed in Figure 1: For each source combination *ij* of feature *i* on a first context dimension and feature *j* on a second dimension, the model has one parameter that describes the probability of recognizing the target as being old (D_{ij}). Furthermore, there is one parameter for recognizing that a distractor is new (D_{New}). This results in 5-item recognition parameters. For each source combination, the parameter d_{ij} represents the probability of the joint retrieval of both source features as compound information (i.e., context binding), resulting in four binding parameters. One source dimension of an item can also be retrieved independently from the other source dimension, represented by two separate parameters (one per source dimension) e_{ij}^{Dim1} and e_{ij}^{Dim2} for each source combination, resulting in eight parameters for the probability of independent source retrieval. The multinomial model also involves guessing parameters allowing for the possibility that participants give the correct answer in the absence of actual memory for an item and its sources. Parameter b is the probability that a participant guesses that an unrecognized item is old. In that case, the model assumes that there is no recollection of the item's source (and therefore also no binding). Thus, if an item is guessed to be old, the sources of the items also have to be guessed. Since guessing the second source dimension could be influenced by the guess of the first dimension, this results in three source guessing parameters in case of guessing that an item is old. Similarly, if the item was recognized as old, participants might have forgotten about one or both of the item's sources and have to guess the respective source information. This also might depend on the assignment of the other source. As can be seen in Figure 1, this results in three source-guessing parameters given that the item was recognized as old.

In MPT modeling, traditionally, response frequencies are aggregated across participants and items assuming that responses by all participants and to all items are independent and identically distributed (i.i.d.; Matzke, Dolan, Batchelder, & Wagenmakers, 2015). However, this assumption rarely holds – even if the pool of participants seems homogeneous (Smith & Batchelder, 2008). However, violations of the i.i.d. assumption can lead to biased parameter estimates, confidence intervals, and statistical tests (Klauer, 2006, 2010; Smith & Batchelder, 2008, 2010).

In hierarchical Bayesian MPT modeling it is assumed that individual parameters arise from a common group-level distribution that is described by hyperparameters. Incorporating heterogeneity into a MPT model has several advantages: Compared to estimating parameters separately for individual participants, additional information for individual parameters is taken from the whole sample, thereby providing more accurate estimates (e.g., Rouder & Lu, 2005). Basically, a hierarchical MPT model consists of a core MPT model with potentially different parameter values for participants specified by a distribution of the individual parameter values (Klauer, 2006). On the individual level, the response probabilities are specified similarly to the model equations in traditional MPT models. However, the hierarchical analysis accounts for the variability across participants by defining a distribution of person parameters on the group level. Different distributional forms for parameter heterogeneity have been suggested, for example, discrete distributions (latent-class approach; Klauer, 2006), beta distributions (beta-MPT; Smith & Batchelder, 2010), or transformed normal distributions (latent-trait approach; Klauer, 2010; Matzke et al., 2015).

We adapted the latent-trait approach (Klauer, 2010) to the multidimensional source-monitoring model proposed by Meiser (2014) using the R package TreeBUGS (Heck, Arnold, & Arnold, 2018). The latent-trait approach assumes that the inverse-probit-transformed MPT parameters for each person follow a multivariate normal distribution on the group level. As explained above, the approach specifies parameters for individual participants as well as parameters on the group level. It also accounts for correlations among the parameters. TreeBUGS uses the program JAGS (Plummer, 2003) to obtain parameter estimates by Markov chain Monte Carlo (MCMC) sampling.

The Current Study

The aim of the study is twofold: First, we want to replicate the result by Meiser et al. (2008) that stochastic

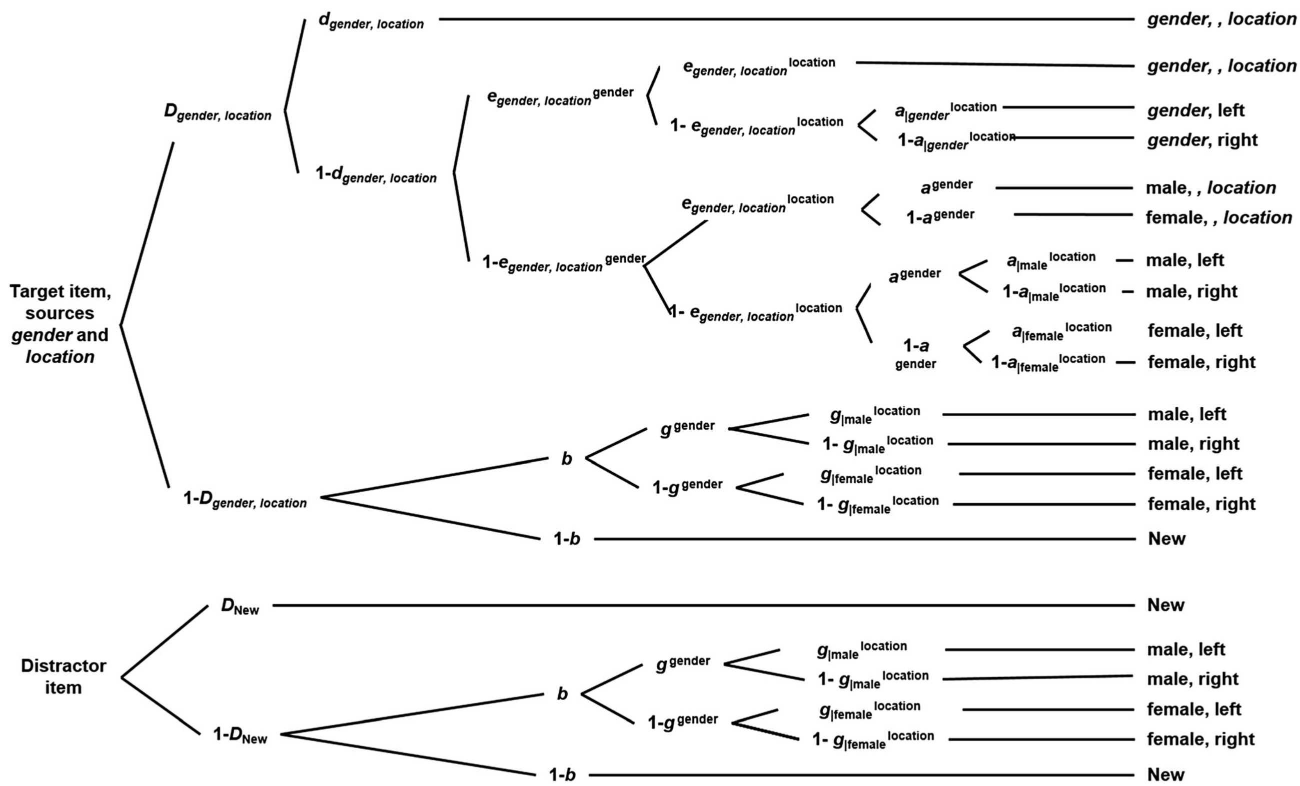

Figure 1. Processing tree diagram of the multinomial model of joint versus independent source memory. $D_{\text{male/female, left/right}}$ = probability of recognizing target items from the sources *gender* and *location*; $d_{\text{male/female, left/right}}$ = probability of joint retrieval of source *gender* on the *gender* dimension and source *location* on the *location* dimension; $e_{\text{male/female, left/right gender}}$ = probability of independent retrieval of source *gender* on the *gender* dimension; $e_{\text{male/female, left/right}}{}^{\text{location}}$ = probability of independent retrieval of source *location* on the *location* dimension; a_{gender} = probability of guessing "male" on the *gender* dimension for recognized target items; $a_{|\text{male}}{}^{\text{location}}$, $a_{|\text{female}}{}^{\text{location}}$ = probability of guessing "left" on the location dimension for recognized target items assigned to male or female, respectively; b = probability of guessing "old"; g^{gender} = probability of guessing "male" on the *gender* dimension for unrecognized target or distractor items; $g_{|\text{male}}{}^{\text{location}}$, $g_{|\text{female}}{}^{\text{location}}$ = probability of guessing "left" on the location *dimension* for unrecognized target or distractor items assigned to male or female, respectively; D_{New} = probability of recognizing distractor items as new. Adapted from "Processing tree diagram of the multinomial model of joint versus independent source memory" by Meiser, 2014, *Experimental Psychology, 61,* p. 408.

dependence in multidimensional source memory is not an artifact due to aggregating across individual respondents. To rule out an artifact of aggregation, we apply a hierarchical version of the MPT model by Meiser (2014). The hierarchical approach includes a latent continuous distribution of memory parameters and thus provides a more appropriate account of person heterogeneity than the analysis by Meiser et al. (2008) who used manifest groups that were defined by observed memory performance (Meiser et al., 2008).[1]

If we achieve the first goal and replicate stochastic dependence with the hierarchical MPT analysis (thereby ruling out an artifact explanation), the second goal is to test predictions about the underlying cognitive processes that lead to stochastic dependency. Here, we employ a manipulation that we expect to influence binding of features but not source memory strength per se. As has been shown by Boywitt and Meiser (2012b) the focus of attention during encoding plays a crucial role for binding in memory. Therefore, we manipulate whether participants focus on joint encoding or separate encoding of the source features. We thus specify an experimental design with two conditions: In the *simultaneous-categorization* condition, participants are asked about conjunctions of both context features for each trial of the learning phase. In the *separate-categorization* condition, participants are asked about one context feature per trial only. If the simultaneous-categorization condition shows larger binding parameters than the separate-categorization condition and if there is no effect or an effect of opposite sign on the source memory parameters *e*, this would provide evidence that context-context

[1] Preliminary evidence was found in a reanalysis of two experiments by Boywitt and Meiser (Experiment 2, 2012a, Experiment 2, 2012b) presented at the Annual Meeting of the Psychonomic Society 2014 (Arnold, Boywitt, & Bröder, 2014). Preliminary results suggested that stochastic dependence is not an artifact of aggregation driven by individual differences in context memory, such that some participants have better memory for both context features than other participants.

binding is possible. Before the experiment, we conducted a simulation study to show that the planned sample size provides sufficiently precise estimates for the model to find such a difference in the binding parameter *d* across the two conditions assuming that such a difference exists (see Simulation Study of Sensitivity section). Thus, the main hypotheses/questions of the study are:

Hypothesis 1 (H1): We expect that, at least in the simultaneous-categorization condition, the binding parameter is larger than zero even if individual differences are controlled for (i.e., a replication of Meiser et al., 2008). This would imply that stochastic dependency is present even when accounting for participant heterogeneity.

Hypothesis 2 (H2): We test whether the simultaneous-categorization condition shows a higher binding parameter *d* than the separate-categorization condition, while the source memory parameters *e* are smaller or equal for the simultaneous-categorization condition compared to the separate-categorization condition. Such a pattern would imply that stochastic dependency of context retrieval can be manipulated separately from independent source memory. As a consequence, the predicted pattern of results between the two experimental conditions – stronger binding in the absence of stronger overall source memory in the simultaneous-categorization condition – would indicate that the stochastic dependence in multidimensional source retrieval is not merely a function of the overall level of source memory across items.

Hypothesis 3 (H3): In addition to the main questions described above, we will test whether binding parameter and source-memory parameters are correlated. If not, this would imply a functional dissociation between binding and independent source recognition.

Method

Participants

To yield reliable parameter estimates (cf. the simulation below), 160 students took part in this study for course credit or monetary compensation. They were randomly assigned to the two conditions with both conditions containing an equal number of participants.

Materials and Design

Eighty words were randomly drawn from a set of 167 concrete German nouns of 4–7 letters (see von Hecker & Meiser, 2005, for details). Additional 40 words were randomly drawn as distractors for the test phase for each participant.

Items were presented either on the left side of the screen or on the right side of the screen. Additionally, they were presented orally either through the left or right speaker of a headphone matching the position on the screen. Items were read either by a female voice (called Christine) or a male voice (called Sebastian). Additionally, a picture of a man or a woman was shown next to the item matching the gender of the voice. Half of the study items appeared on the left side of the screen, and the other half appear on the right side of the screen. Half of the items on either side were presented by the female voice and half by the male voice. The study phase was divided into 5 blocks of 16 items within each block, each combination of the feature's location and gender appeared four times in random order within one block. The blocks were not noticeable for participants and the assignment of words to context combinations was randomized for each participant.

Procedure

All instructions and materials were presented via computer. After being introduced to the speakers Christine and Sebastian by listening to them reading a short passage of a German translation of Winnie-The-Pooh (Milne, 1973), participants were told that they would have to learn a series of words presented randomly by the two speakers. There were no explicit source-monitoring instructions. In each trial of the learning phase, a word was presented for 3 s and participants had to answer a question concerning the gender of speaker and the location of the word on the screen after word presentation. In the *separate-categorization* condition, participants were asked either about the gender of the speaker (i.e., whether the word was presented by Christine OR by Sebastian) or about the location of the item (i.e., whether the word was presented left OR right). Both questions were asked equally often but the order of the questions across items was randomized for each participant. In the *simultaneous-categorization* condition, participants were asked about gender and location simultaneously at each trial. For example, they were asked whether the item had been presented by Christine on the left side/Sebastian on the right side OR by Sebastian on the left side/Christine on the right side. In both conditions, participants had to choose between two options by pressing either the *F* key or the *J* key. Then, the next item was presented.

After the study phase, participants had to solve Raven matrices for three minutes as a distractor task. Afterward, the instructions for the memory test were given. Participants were asked to judge whether an item was old or new. If they judged it as old, they were asked to indicate gender and location of the item by choosing between four response fields each reflecting a possible combination of gender and location. To avoid possible confusions, the options "Christine, left" and "Sebastian, left" were both presented on the left side of the screen, while the two "right"-options were presented on the right side of the screen. The test phase consisted of the 80 target items and 40 distractors displayed in random order in the center of the screen. All responses to the memory test were given via mouse clicking.

Analyses

Model-based analyses used Bayesian estimation. For all other (i.e., frequentist) analyses, we set the conventional significance level to $\alpha = 5\%$. We planned to exclude participants showing a higher or equal false-alarm rate than hit rate.

Overall Recognition and Context Memory

For illustrative purposes and following customary analyses in source memory studies, memory performance is reported in terms of *Hit Rate* (i.e., probability of correctly identifying an old item as "old") – *False Alarms Rate* (i.e., probability of falsely calling a new item "old") and the average conditional source identification measure for both source dimensions (ACSIM; Murnane & Bayen, 1996). Both measures are compared between the two groups via a two-sample *t*-test.

Model-Based Analyses

Since our hypotheses refer to the MPT model parameters, most of the analyses are model-based and compare model parameters between the groups. The model as described above and as shown in Figure 1 is not identifiable. Therefore, we imposed theoretically based restrictions: Item recognition is assumed to be independent from the presenting source resulting in one item-recognition parameter D regardless of the source. Following the usual procedure, distractor detection has been set equal to the item-recognition parameter (cf. Meiser & Bröder, 2002). Additionally, binding is assumed to be independent of source combination, resulting in one overall binding probability d. Independent recognition of the sources is assumed to rely only on the source dimensions resulting in two source-recognition parameters e^{gender} and e^{location}. Parameter b represents the probability that a participant guesses that an unrecognized item is old. Source guessing is assumed to be independent of whether the item was recognized or guessed as old before. Guessing the source of the second dimension can still be influenced by the response on the first dimension, so that three source-guessing parameters g_{gender}, $g_{\text{location}}|\text{male}$, and $g_{\text{location}}|\text{female}$ are estimated. These restrictions have been used in many previous applications of this paradigm (e.g., Boywitt & Meiser, 2012a, 2012b) and in similar source monitoring paradigms (e.g., Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013). The resulting model equations and parameter restrictions can be found in Appendix A. The resulting R code can be found at https://osf.io/kw3pv/. The model was fitted to the two conditions separately thus estimating different mean vectors and covariance matrices for the two hierarchical distributions. Based on the posterior samples, a credibility interval of the difference between group specific parameters (cf. Smith & Batchelder, 2010) was computed. The main hypothesis was assessed by checking whether the 95% Bayesian credibility interval (95% BCI) does not include zero, thereby providing evidence for a difference between the two groups. Additionally, TreeBUGS computes Bayesian *p*-values for the test (Bayesian *p*-values are defined as the proportion of posterior samples of a parameter difference that are below or above zero).

For H1, we tested whether the binding parameters of the two conditions differ from zero. If the 95% BCI contains zero, this indicates the absence of binding in the respective condition. If for at least one condition, the BCI of the binding parameter does not include zero, this is a replication of Meiser et al. (2008) and Arnold, Boywitt, and Bröder (2014) and shows that binding is not an artifact of aggregation over participants.

For H2, we compared the binding parameters d and the source recognition parameters e^{gender} and e^{location} between the two conditions. If the 95% BCIs overlap, the conditions do not differ substantially with regard to that process. If d is larger in the simultaneous-categorization condition relative to the separate-categorization condition, this would show that binding depends on the orienting task during encoding and allows us to test whether the effect of binding can be dissociated from an effect on the source-memory parameters e^{gender} and e^{location}. Importantly, a selective increase in d but not in e^{gender} and e^{location} would support the view that binding is a cognitive process in its own right that can be dissociated from the overall level of source memory.

H3: Correlations between the person parameters are implemented as free parameters in the latent-trait approach and are thus estimated automatically. We applied a similar logic to the hypothesis testing described above and examined whether the 95% BCI includes zero. For the correlation between binding and source-recognition parameters, we expected the 95% BCI to include zero if binding and source recognition can be functionally dissociated.

All MPT analyses were carried out with TreeBUGS (Heck, Arnold, & Arnold, 2018) with 100,000 iterations and the first 20,000 discarded as burn-in period.

Table 1. Probability-transformed group-level parameters $\Phi(\mu)$ of the multidimensional source-monitoring model of memory for multidimensional source information in the sensitivity simulation

	Separate condition				Simultaneous condition			
Parameter	True	Mean (*M*)	*SD* (*M*)	Mean (*S*)	True	Mean (*M*)	*SD* (*M*)	Mean (*S*)
D	.60	0.60	0.03	.03	.60	0.60	0.03	.03
d	**.00**	0.02	0.01	.01	**.15**	0.14	0.03	.03
e_1	.30	0.27	0.05	.05	.30	0.30	0.05	.06
e_2	.10	0.08	0.03	.03	.10	0.11	0.04	.04
b	.30	0.30	0.03	.03	.30	0.30	0.03	.03
a_1	.50	0.50	0.01	.01	.50	0.50	0.01	.01
a_{21}	.50	0.50	0.01	.01	.50	0.50	0.02	.02
a_{22}	.50	0.50	0.01	.01	.50	0.50	0.02	.02

Note. Posterior mean and standard deviation of the inverse-probit transformed group-level parameters $\Phi(\mu)$ per replication are labeled by "*M*" and "*S*", respectively. The summary statistics mean () and *SD* () are the sample means and standard deviations of the estimates across 750 replications of the sensitivity simulation in which a true difference in the binding parameter *d* across conditions is assumed (i.e., d^{diff} = .15 as highlighted in bold).

By default, TreeBUGS only retains every fifth iteration for computing summary statistics to reduce autocorrelation. To assess goodness of fit, we also calculated posterior-predictive tests (Meng, 1994) and the test statistics T_1 and T_2 (Klauer, 2010) that focus on the mean and covariance of the individual frequencies, respectively.

Simulation Study of Sensitivity

To assess the expected precision of the parameter estimates for the hierarchical model of memory for multidimensional source information, we performed a Monte Carlo simulation. Similar to a power analysis in the frequentist framework, this simulation provides a priori estimates how precisely the parameters can be estimated within the Bayesian framework for a given sample size. The simulations were performed using the data-generation and analysis functions provided in the R package TreeBUGS (Heck, Arnold, & Arnold, 2018). Note that the simulation results were tailored to the present model and design and do not necessarily generalize to other scenarios.

Following the preregistration protocol, we simulated 1,500 replications with $n_1 = n_2 = 80$ participants per condition, each providing 120 responses (20 per source combination and 40 distractor items). To test the sensitivity of the hierarchical model, data were generated under two scenarios assuming (a) a larger binding parameters *d* in the simultaneous-categorization than in the separate-categorization condition (i.e., $d^{\text{sim}} = .15$ and $d^{\text{sep}} = .00$, resulting in $d^{\text{diff}} = .15$) and (b) identical binding parameters *d* across conditions (i.e., $d^{\text{sim}} = d^{\text{sep}} = .00$, resulting in $d^{\text{diff}} = .00$). The remaining data-generating parameters were identical in the two experimental conditions and selected based on the parameter estimates of a reanalysis of multidimensional source-monitoring data sets (Boywitt & Meiser, 2012a, 2012b). The group-level means $\Phi(\mu)$ are shown in Table 1 and reflect medium item memory ($D = .60$) and that the sources can be independently encoded to a different degree ($e_1 = .30$ vs. $e_2 = .10$). We assumed independence of the parameters by specifying a diagonal covariance matrix Σ. However, based on previous empirical estimates, different standard deviations σ on the probit scale were chosen for the parameters, thereby assuming that the hypothesized cognitive processes vary across persons to different degrees ($\sigma_{\text{d}} = \sigma_{\text{a1}} = \sigma_{\text{a21}} = \sigma_{\text{a22}} = 0.20$; $\sigma_{\text{D}} = \sigma_{\text{b}} = 0.60$; $\sigma_{\text{e1}} = \sigma_{\text{e2}} = 0.80$).

Table 1 summarizes the parameter estimates for the subset of 750 replications assuming a true difference in the binding parameter across conditions (i.e., $d^{\text{diff}} = .15$). As expected, the parameters are recovered without bias, that is, the means of the estimated parameters across replications match the data-generating parameters. Moreover, the standard deviation of the mean estimates across replications is in line with the posterior standard deviation, a summary statistic of the posterior distribution that is often used to quantify estimation uncertainty. Hence, posterior means and posterior standard deviations provide unbiased estimates. Moreover, the posterior standard deviations are sufficiently small to allow for substantive interpretations of the parameters. This shows that the chosen sample size of $n_1 = n_2 = 80$ participants per condition ensures sufficiently precise estimates.

The main interest is in the comparison of the binding parameter *d* across the two experimental conditions. Hence, we computed the difference of the posterior samples of $\Phi(\mu_d)$ across the two conditions within each replication. Thereby, we can test whether the model can reliably detect the presence or absence of an effect of the simultaneous presentation of stimuli as assumed by the two simulation scenarios (i.e., $d^{\text{diff}} = .15$ and $d^{\text{diff}} = .00$). Figure 2 shows that this was indeed the case, and that the estimated differences allowed for a reliable detection of the effect of presentation format on the binding parameter *d* across the two conditions. Moreover, the 95% BCI included zero

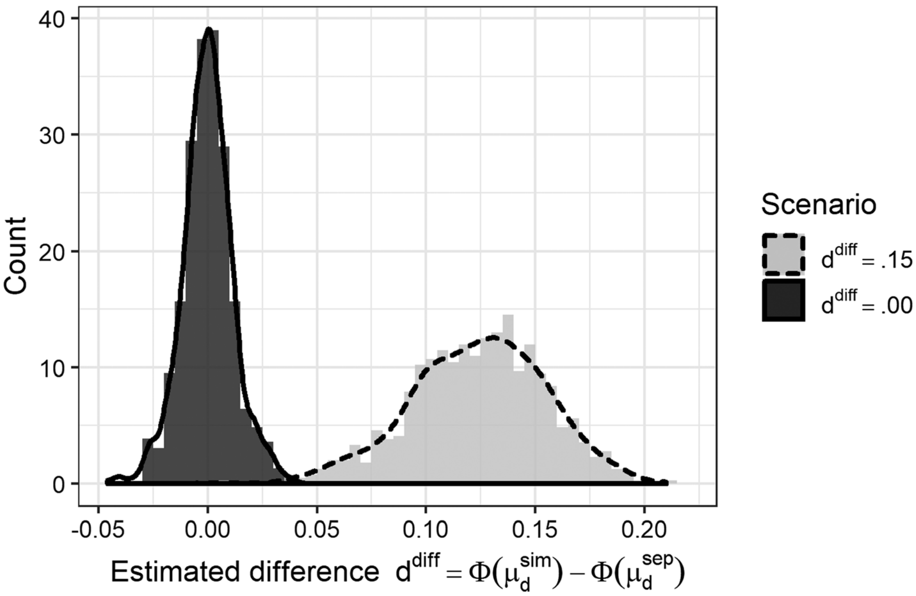

Figure 2. Distribution of the estimated difference d^{diff} in the binding parameter d for the two conditions across 1,500 replications of the sensitivity simulation.

in 100% of the replications in which an effect on d was absent (i.e., if $d^{diff} = .00$), and excluded zero in 91.7% of the replications in which an effect on d was present (i.e., if $d^{diff} = .15$). Hence, for the given sample size, number of responses, and data-generating parameters, the Bayesian hierarchical model of memory for multidimensional source information provides a reliable test whether simultaneous presentation of stimuli results in context binding.

Informative Data Patterns

The crucial test regarding an artifact of aggregation over participants is H1. If we find a binding parameter that is different from zero, stochastic dependency is not an artifact of aggregation. This would replicate the findings by Meiser et al. (2008) and Arnold et al. (2014). Additionally, this is a precondition for the second goal of the study: testing different accounts for stochastic dependency.

H2 reflects the test regarding the explanations of stochastic dependency. If there is no difference in the independent source memory parameters e^{gender} and $e^{location}$ between conditions, but the binding parameter d is higher in the simultaneous-categorization condition, then the change of the binding parameter cannot be attributed to an unspecific overall increase in source memory across items. This pattern would thus speak in favor of context-context binding. It might also be possible that the manipulations affect working memory and therefore participants in the *simultaneous-monitoring* condition show worse memory performance than participants in the *one-dimension-monitoring* condition. In contrast, no group difference in the binding parameter speaks in favor of item-context binding.

The absence of a correlation between binding parameters d and source memory parameters e (H3) reflects the assumption of functional independence of binding and source memory.

Results

No participant showed a higher or equal false alarm than hit rate. Thus, no data were excluded. However, due to an error in the assignment of participants to groups, group sizes were not perfectly balanced between conditions, leading to $n = 83$ participants in the *separate-categorization* condition (61 female, $M_{age} = 24.20$, $SD = 5.57$), and $n = 77$ participants in the *simultaneous-categorization* condition (62 female, $M_{age} = 25.21$, $SD = 8.44$). Participants were recruited at the Universities of Mannheim and Heidelberg. They received course credit or monetary compensation.

Overall Recognition and Context Memory

Table 2 shows the difference between the overall *Hit Rate* and *False Alarm Rate* and the ACSIM (defined as the average proportion of correct source judgments for all items judged as old; Murnane & Bayen, 1996) for both conditions. With respect to the difference between the *Hit Rate* and *False Alarm Rate*, the two conditions did not differ significantly, $t(158) = 0.36$, $p = .72$. Similarly, for ACSIM, the two conditions did not differ significantly, $t(158) = .87$, $p = .39$.

Table 2. Descriptive measures of source memory

	Separate condition		Simultaneous condition	
	M	*SD*	*M*	*SD*
Difference: Hit Rate – False-Alarm Rate	0.41	0.18	0.42	0.20
ACSIM	0.23	0.11	0.25	0.11

Notes. *M* = Mean and *SD* = standard deviation across the individual estimates of the corresponding statistics per participant. ACSIM is defined as average conditional source identification performance (CSIM) for all four source combinations For the calculation of CSIM, the number of items from a source (e.g., female left) that was attributed to the correct source is divided by the number of item hits for items from this source (i.e., all items from this source for which a source judgment was given).

Table 3. Parameter estimates of the multidimensional source-monitoring model

	Separate Condition		Simultaneous Condition	
Parameter	Mean	95% BCI	Mean	95% BCI
D	.39	[.35, .44]	.40	[.35, .45]
d	.14	[.07, .23]	.19	[.10, .27]
e_1	.29	[.19, .39]	.27	[.18, .35]
e_2	.18	[.07, .29]	.08	[.01, .17]
b	.14	[.11, .18]	.21	[.16, .26]
a_1	.57	[.52, .62]	.55	[.50, .60]
a_{21}	.59	[.52, .68]	.47	[.40, .53]
a_{22}	.43	[.33, .52]	.50	[.42, .58]

Note. Posterior mean and 95% Bayesian credibility intervals of the probability-transformed group-level parameters $M = \Phi(\mu)$.

Model-Based Analyses

As described above, the model was fitted to the two conditions separately thus allowing for separate hierarchical group-level distributions per condition. For all parameters, the potential scale reduction factor was $\hat{R} < 1.06$, indicating good convergence. The goodness-of-fit statistics T_1 and T_2 (Klauer, 2010) test whether the model predictions are in line with the observed mean frequencies and covariance across participants, respectively. The corresponding posterior-predictive *p*-values were $p_1 = .46$ and $p_2 = .40$ in the simultaneous condition and $p_1 = .10$ and $p_2 = .08$ in the separate condition, thus indicating a good model fit. Moreover, Table 3 shows the group-level MPT parameters on the probability scale for both conditions.

To test H1, we checked whether the binding parameters of the two conditions differed from zero. For the *separate-categorization* condition, the group-level binding parameter had a posterior mean and standard deviation of $M = .14$ and $SD = .05$, respectively, with a 95% BCI of [.07, .23]. For the *simultaneous-categorization* condition, the binding parameter on the group level was estimated to be $M = .19$ and $SD = .04$. The 95% BCI was [.10, .27]. This indicates the presence of binding in both conditions. Thereby, we replicated the results of Meiser et al. (2008) and Arnold et al. (2014) while accounting for between-subject heterogeneity.

To test H2, we compared the binding parameters d and the source recognition parameters e^{gender} and e^{location} between the two conditions. The posterior distributions of the group-level parameters on the probability scale for both conditions are shown in Figure 3. As described above, for the binding parameters d, the 95% BCIs for the two conditions overlapped. Therefore, the conditions did not differ regarding binding. We tested the main hypothesis by checking whether the 95% BCI of the difference of the binding parameter between the two conditions ($d_{\text{sep}} - d_{\text{sim}}$) included zero. The estimated credibility interval was [−.18, .07], thus providing evidence against a difference between the two conditions. In line with this result, the corresponding Bayesian *p*-value in favor of a larger d parameter in the simultaneous condition was $p_B = .80$. For the *separate-categorization* condition, the source-recognition parameters e^{gender} on the group level had a posterior mean and standard deviation of $M = .29$ and $SD = .05$, respectively, with a 95% BCI of [.19, .39]; the estimates for the source-recognition parameters e^{location} were $M = .18$ and $SD = .06$ with a 95% BCI of [.07, .29]. For the *simultaneous-categorization* condition, the source-recognition parameters e^{gender} on the group level was $M = .27$ with $SD = .04$ with a 95% BCI of [.18, .35]; and the estimates of the source-recognition parameters e^{location} were $M = .08$ and $SD = .05$ with a 95% BCI of [.01, .17]. Again, the BCIs for the source-recognition parameters overlapped. In line with these results, the corresponding Bayesian *p*-values were $p_B = .44$ for a larger e^{gender} parameter in the simultaneous condition, and $p_B = .12$ for a larger e^{location} parameter in the simultaneous condition. Hence, the conditions did not differ with regard to source recognition.

To test H3, we checked whether the 95% BCI for the correlation between binding and source-recognition parameters included zero. For the *separate-categorization* condition, the correlation between the binding parameter and the source-recognition parameter e^{gender} had a posterior mean and standard deviation of $M = -.01$ and $SD = .36$ with a 95% BCI of [−.65, .66]; the correlation between the binding parameter and the source recognition parameter e^{location} was distributed with $M = .01$ and $SD = .36$. The 95% BCI was [−.63, .68]. For the *simultaneous-categorization* condition, the mean correlation between the binding parameter and the source recognition parameters e^{gender} was $M = .08$ with $SD = .28$. The 95% BCI was [−.48, .60]; the correlation between the binding parameter and the source recognition parameter e^{location} was $M = -.14$ with $SD = .30$. The 95% BCI was [−.70, .44]. As expected, this indicates that binding and source recognition can be functionally dissociated. However, given the large credibility

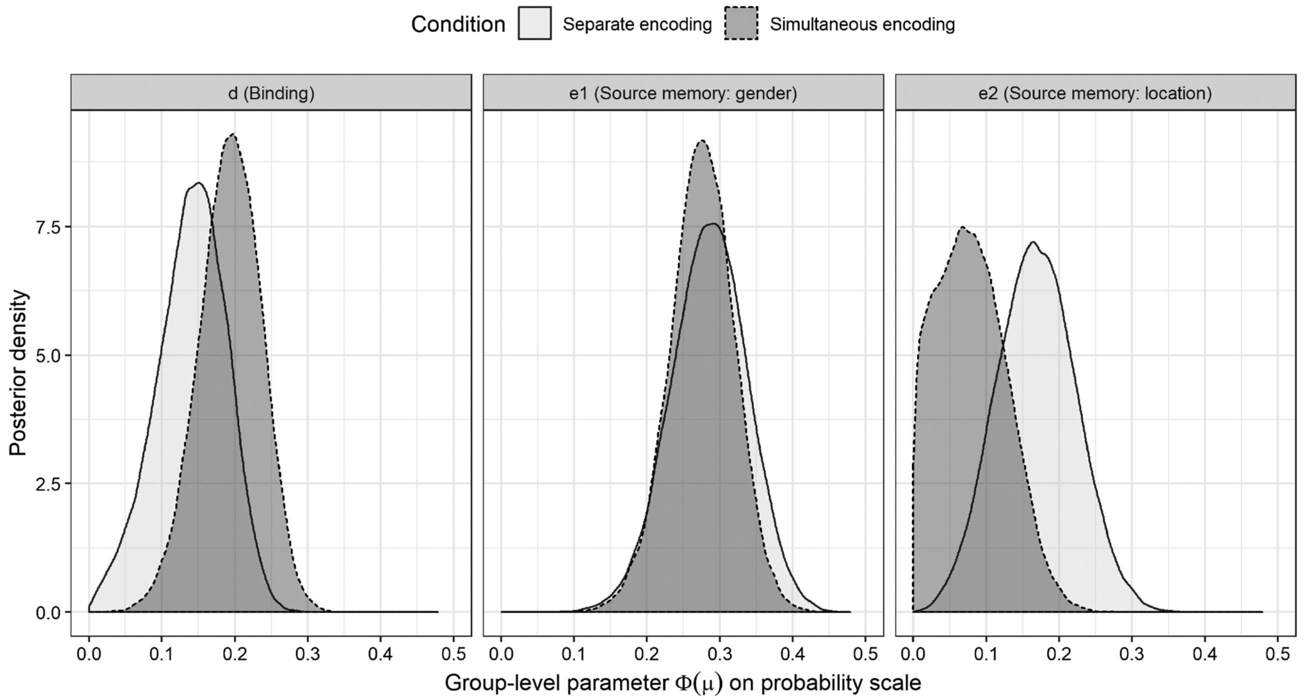

Figure 3. Posterior distributions of the inverse-probit transformed group-level parameters $\Phi(\mu)$ on the probability scale for the binding parameter d and the source-memory parameters $e^{\text{gender}} = e_1$ and $e^{\text{location}} = e_2$ in the encoding conditions.

intervals of the correlation coefficients, the present results do not provide strong evidence for the absence of an effect.

For completeness, we also fitted the standard, non-hierarchical source-monitoring model using maximum-likelihood estimation (Meiser, 2014). In Appendix B, we show that this MPT model had a good fit, resulted in very similar point estimates for the parameters, and lead to identical substantive conclusions.

Discussion

In both conditions, we found binding parameters that were larger than zero (H1) while accounting for heterogeneity in source memory by fitting a hierarchical MPT model. Therefore, we replicated the findings by Meiser et al. (2008) and Arnold et al. (2014) and showed that stochastic dependency in source memory is not merely an artifact of the aggregation of response frequencies over participants.

Across the simultaneous and separate encoding conditions, we did neither find a difference in the binding parameter d (H2) nor in the independent source-memory parameters e^{gender} and e^{location}. The absence of a group difference in the binding parameter speaks in favor of item-context binding. Note, however, that our manipulation of the context encoding strategies was rather ad hoc. Although it was based on results showing that an attentional focus is necessary for binding (Boywitt & Meiser 2012a), the procedure might not have been efficient for inducing different context encoding strategies in both conditions.

The absence of a correlation between binding parameters d and source memory parameters e (H3) reflects the assumption of functional independence of binding and source memory. However, given the large credibility intervals of the correlation, the current study could only provide weak evidence for this conclusion. Note that the relatively large posterior uncertainty of the correlation parameter can be attributed to the discrete nature of the data (which provide less information compared to continuous data) and the definition of the binding and source-memory parameters as conditional probabilities (implying that these parameters can only be estimated reliably if the item-memory parameter D is large).

Taken together, the present study yielded mixed results. The experiment replicated the earlier finding of stochastic dependency in multidimensional source retrieval while controlling for interindividual differences. Stochastic dependency has been interpreted in terms of binding across context features in multidimensional source memory (e.g., Meiser & Bröder, 2002), and the current hierarchical MPT analysis ruled out an artifact interpretation of stochastic dependence as spurious correlation due to heterogeneity of source memory performance across participants. The experimental manipulation of attentional focus between the simultaneous-categorization and separate-categorization condition, however, did neither exert an effect on the binding parameter nor on the parameters of independent source memory. An effect on the binding parameter was expected from the theoretical perspective that stochastic dependency is driven by binding among context features during encoding, whereas the item-context hypothesis did not imply a difference in binding between the categorization conditions. As a consequence, the current findings do not reveal unequivocal evidence to distinguish between the context-context binding account (Boywitt & Meiser, 2012a, 2012b; Meiser & Bröder, 2002) and the item-context binding account (Hicks & Starns, 2015; Starns & Hicks, 2005, 2008) of the observed stochastic relation in multidimensional source retrieval.

References

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 & Review, 20*, 326–333. https://doi.org/10.3758/s13423-012-0342-7

Arnold, N. R., Boywitt, C. D., & Bröder, A. (2014, November). *Context memory and context binding: A hierarchical multinomial modeling approach*. Poster at the 5th Annual Meeting of the Psychonomic Society (Psychonomics), Long Beach, CA.

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. https://doi.org/10.1037/0278-7393.22.1.197

Boywitt, C. D., & Meiser, T. (2012a). Bound context features are integrated at encoding. *The Quarterly Journal of Experimental Psychology, 65*, 1484–1501. https://doi.org/10.1080/17470218.2012.656668

Boywitt, C. D., & Meiser, T. (2012b). The role of attention for context-context binding of intrinsic and extrinsic features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1099–1107. https://doi.org/10.1037/a0026988

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. https://doi.org/10.3758/s13428-017-0869-7

Hicks, J. L., & Starns, J. J. (2015). Using multidimensional encoding and retrieval contexts to enhance our understanding of stochastic dependence in source memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 62, pp. 101–140). Cambridge, MA; Academic Press. https://doi.org/10.1016/bs.plm.2014.09.004

Hicks, J. L., & Starns, J. J. (2016). Successful cuing of gender source memory does not improve location source memory. *Memory & Cognition, 44*, 650–659. https://doi.org/10.3758/s13421-016-0586-y

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

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71*, 7–31. https://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. https://doi.org/10.1007/s11336-009-9141-0

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. https://doi.org/10.1007/s11336-013-9374-9

Meiser, T. (2014). Analyzing stochastic dependence of cognitive processes in multidimensional source recognition. *Experimental Psychology, 61*, 402–415. https://doi.org/10.1027/1618-3169/a000261

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

Meiser, T., Sattler, C., & Weisser, K. (2008). Binding of multidimensional context information as a distinctive characteristic of remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 32–49. https://doi.org/10.1037/0278-7393.34.1.32

Meng, X.-L. (1994). Posterior predictive *p*-values. *The Annals of Statistics, 22*, 1142–1160. https://doi.org/10.1214/aos/1176325622

Milne, A. A. (1973). *Winnie-the-pooh*. London, UK: Methuen Children's Books.

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

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition, 24*, 417–428. https://doi.org/10.3758/BF03200931

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In M. Plummer, K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Com puting* (Vol. 124, pp. 125–134). Vienna, Austria: Technical University.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review, 12*, 573–604. https://doi.org/10.3758/BF03196750

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review, 15*, 713–731. https://doi.org/10.3758/PBR.15.4.713

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

Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1213–1220. https://doi.org/10.1037/0278-7393.31.6.1213

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review, 15*, 309–314. https://doi.org/10.3758/PBR.15.2.309

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12. https://doi.org/10.1037/h0080017

Vogt, V., & Bröder, A. (2007). Independent retrieval of source dimensions: An extension of results by Starns and Hicks (2005) and a comment on the ACSIM measure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 443–450. https://doi.org/10.1037/0278-7393.33.2.443

von Hecker, U., & Meiser, T. (2005). Defocused attention in depressed mood: Evidence from source monitoring. *Emotion, 5*, 456–463. https://doi.org/10.1037/1528-3542.5.4.456

History
Received March 10, 2017
Revision received January 26, 2019
Accepted January 28, 2019
Published online May 17, 2019

Conflict of Interest
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Open Data
The data and code can be accessed via the Open Science Framework at https://osf.io/kw3pv/.

ORCID
Nina Rebecca Arnold
https://orcid.org/0000-0002-9442-7122

Nina Rebecca Arnold
Department of Psychology
School of Social Sciences
University of Mannheim
L13 17
68161 Mannheim
Germany
arnold@uni-mannheim.de

Appendix A

Resulting Model Equations and Parameter Restrictions

####### MPT model equations (eqn) and restrictions

#eqn file

AA	AA_AA	D11*d_11
AA	AA_AA	D11*(1-d_11)*e1_11*e2_11
AA	AA_AA	D11*(1-d_11)*e1_11*(1-e2_11)*a21
AA	AA_AB	D11*(1-d_11)*e1_11*(1-e2_11)*(1-a21)
AA	AA_AA	D11*(1-d_11)*(1-e1_11)*e2_11*a1
AA	AA_BA	D11*(1-d_11)*(1-e1_11)*e2_11*(1-a1)
AA	AA_AA	D11*(1-d_11)*(1-e1_11)*(1-e2_11)*a1*a21
AA	AA_AB	D11*(1-d_11)*(1-e1_11)*(1-e2_11)*a1*(1-a21)
AA	AA_BA	D11*(1-d_11)*(1-e1_11)*(1-e2_11)*(1-a1)*a22
AA	AA_BB	D11*(1-d_11)*(1-e1_11)*(1-e2_11)*(1-a1)*(1-a22)
AA	AA_AA	(1-D11)*b*g1*g21
AA	AA_AB	(1-D11)*b*g1*(1-g21)
AA	AA_BA	(1-D11)*b*(1-g1)*g22
AA	AA_BB	(1-D11)*b*(1-g1)*(1-g22)
AA	AA_N	(1-D11)*(1-b)
AB	AB_AB	D12*d_12
AB	AB_AB	D12*(1-d_12)*e1_12*e2_12
AB	AB_AA	D12*(1-d_12)*e1_12*(1-e2_12)*a21
AB	AB_AB	D12*(1-d_12)*e1_12*(1-e2_12)*(1-a21)
AB	AB_AB	D12*(1-d_12)*(1-e1_12)*e2_12*a1
AB	AB_BB	D12*(1-d_12)*(1-e1_12)*e2_12*(1-a1)
AB	AB_AA	D12*(1-d_12)*(1-e1_12)*(1-e2_12)*a1*a21
AB	AB_AB	D12*(1-d_12)*(1-e1_12)*(1-e2_12)*a1*(1-a21)
AB	AB_BA	D12*(1-d_12)*(1-e1_12)*(1-e2_12)*(1-a1)*a22
AB	AB_BB	D12*(1-d_12)*(1-e1_12)*(1-e2_12)*(1-a1)*(1-a22)
AB	AB_AA	(1-D12)*b*g1*g21
AB	AB_AB	(1-D12)*b*g1*(1-g21)
AB	AB_BA	(1-D12)*b*(1-g1)*g22
AB	AB_BB	(1-D12)*b*(1-g1)*(1-g22)
AB	AB_N	(1-D12)*(1-b)
BA	BA_BA	D21*d_21
BA	BA_BA	D21*(1-d_21)*e1_21*e2_21
BA	BA_BA	D21*(1-d_21)*e1_21*(1-e2_21)*a22
BA	BA_BB	D21*(1-d_21)*e1_21*(1-e2_21)*(1-a22)
BA	BA_AA	D21*(1-d_21)*(1-e1_21)*e2_21*a1
BA	BA_BA	D21*(1-d_21)*(1-e1_21)*e2_21*(1-a1)
BA	BA_AA	D21*(1-d_21)*(1-e1_21)*(1-e2_21)*a1*a21
BA	BA_AB	D21*(1-d_21)*(1-e1_21)*(1-e2_21)*a1*(1-a21)
BA	BA_BA	D21*(1-d_21)*(1-e1_21)*(1-e2_21)*(1-a1)*a22
BA	BA_BB	D21*(1-d_21)*(1-e1_21)*(1-e2_21)*(1-a1)*(1-a22)
BA	BA_AA	(1-D21)*b*g1*g21
BA	BA_AB	(1-D21)*b*g1*(1-g21)
BA	BA_BA	(1-D21)*b*(1-g1)*g22
BA	BA_BB	(1-D21)*b*(1-g1)*(1-g22)
BA	BA_N	(1-D21)*(1-b)
BB	BB_BB	D22*d_22
BB	BB_BB	D22*(1-d_22)*e1_22*e2_22
BB	BB_BA	D22*(1-d_22)*e1_22*(1-e2_22)*a22
BB	BB_BB	D22*(1-d_22)*e1_22*(1-e2_22)*(1-a22)
BB	BB_AB	D22*(1-d_22)*(1-e1_22)*e2_22*a1
BB	BB_BB	D22*(1-d_22)*(1-e1_22)*e2_22*(1-a1)
BB	BB_AA	D22*(1-d_22)*(1-e1_22)*(1-e2_22)*a1*a21
BB	BB_AB	D22*(1-d_22)*(1-e1_22)*(1-e2_22)*a1*(1-a21)
BB	BB_BA	D22*(1-d_22)*(1-e1_22)*(1-e2_22)*(1-a1)*a22
BB	BB_BB	D22*(1-d_22)*(1-e1_22)*(1-e2_22)*(1-a1)*(1-a22)
BB	BB_AA	(1-D22)*b*g1*g21
BB	BB_AB	(1-D22)*b*g1*(1-g21)
BB	BB_BA	(1-D22)*b*(1-g1)*g22
BB	BB_BB	(1-D22)*b*(1-g1)*(1-g22)
BB	BB_N	(1-D22)*(1-b)
N	N_N	DN
N	N_AA	(1-DN)*b*g1*g21
N	N_AB	(1-DN)*b*g1*(1-g21)
N	N_BA	(1-DN)*b*(1-g1)*g22

```
N    N_BB    (1-DN)*b*(1-g1)*(1-g22)
N    N_N     (1-DN)*(1-b)

#restrictions file
"D11 = D12 = D21 = D22 = DN",
"d_11 = d_12 = d_21 = d_22",
"e1_11 = e1_12 = e1_21 = e1_22",
"e2_11 = e2_12 = e2_21 = e2_22",
"g1 = a1",
"g21 = a21",
"g22 = a22"
```

Appendix B

Table B1. Comparison of non-hierarchical and hierarchical parameter estimates

	Separate condition				Simultaneous condition			
	Non-hierarchical analysis		Hierarchical analysis		Non-hierarchical analysis		Hierarchical analysis	
Parameter	Estimate	95% CI	Mean (*M*)	95% BCI	Estimate	95% CI	Mean (*M*)	95% BCI
D	.41	[.39, .42]	.39	[.35, .44]	.42	[.40, .44]	.40	[.35, .45]
d	**.23**	**[.17, .29]**	.14	[.07, .23]	**.22**	**[.16, .29]**	.19	[.10, .27]
e_1	.26	[.19, .34]	.29	[.19, .39]	.29	[.20, .37]	.27	[.18, .35]
e_2	.16	[.07, .24]	.18	[.07, .29]	.13	[.03, .22]	.08	[.01, .17]
b	.18	[.16, .19]	.14	[.11, .18]	.26	[.25, .28]	.21	[.16, .26]
a_1	.54	[.51, .56]	.57	[.52, .62]	.52	[.50, .54]	.55	[.50, .60]
a_{21}	.55	[.52, .58]	.59	[.52, .68]	.48	[.46, .51]	.47	[.40, .53]
a_{22}	.47	[.44, .50]	.43	[.33, .52]	.49	[.46, .51]	.50	[.42, .58]

Note. For the non-hierarchical analysis we used the program multiTree (Moshagen, 2010) and we report maximum likelihood estimates and confidence intervals. Both conditions were estimated separately and showed good model-fit: $G^2_{\text{separate}}(12) = 12.20$, $p = .43$; $G^2_{\text{simultaneous}}(12) = 10.97$, $p = .53$. For the hierarchical analysis, we present posterior means and 95% Bayesian credibility intervals of group-level medians as reported in the main text.