Journal of Memory and Language 76 (2014) 216–236

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Modeling source-memory overdistribution

David Kellen *, Henrik Singmann, Karl Christoph Klauer

Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, D-79085 Freiburg, Germany

ARTICLE INFO

Article history:
Received 30 January 2014
revision received 27 June 2014
Available online 9 August 2014

Keywords:
Overdistribution
Source memory
Familiarity
Recollection
MPT modeling
Hierarchical-Bayesian modeling

ABSTRACT

In a process-dissociation task of source memory, individuals have to judge whether items belong to one of different, mutually exclusive contexts (e.g., Source A, Source B). The acceptance rates to different test probes (e.g., "Source A?") can be used to estimate the probability that the item is assigned simultaneously to the different contexts ("Source A *and* Source B"), designated as *source overdistribution*. Brainerd et al. (2012) have argued that source overdistribution can be used to refute traditional models of source memory such as the One or Two High-Threshold Source-Memory models (1HTSM and 2HTSM; Batchelder and Riefer, 1990; Bayen et al., 1996). We reanalyze previously-published datasets, including Brainerd et al.'s data, and show that there is no support for the rejection of the 1HTSM/2HTSM. Moreover, through a hierarchical-Bayesian model comparison using data from two new experiments, we show that the 2HTSM is not only able to account for source overdistribution, but also provides the best account of the data among different candidate models. These new results suggest that source overdistribution is an outcome of different guessing processes.

Introduction

An important distinction in the memory literature is between *item memory* and *source memory* (Johnson, Hashtroudi, & Lindsay, 1993). While item memory concerns the ability to remember previously acquired information (e.g., "Did I see this word before in the experiment?"), source memory is concerned with contextual details associated with the acquisition of information (e.g., "who said this word?"). The relationship between these two types of memory has produced a considerable body of work along with a diverse set of models (Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996; Hautus, Macmillan, & Rotello, 2008; Klauer & Kellen, 2010; Meiser & Bröder, 2002; Onyper, Zhang, & Howard, 2010; Qin, Raye, Johnson, & Mitchell, 2001; Schütz & Bröder, 2011).

Despite some divergences in the literature, there is a considerable level of convergence regarding the relationship between item and source memory (Klauer & Kellen, 2010; Onyper et al., 2010). This understanding of item and source memory has recently been questioned by Brainerd, Reyna, Holliday, and Nakamura (2012), who reported a phenomenon entitled *source overdistribution* that they argued to be incompatible with current modeling approaches, such as the One and Two-High Threshold Source-Memory Models (1HTSM and 2HTSM; Batchelder & Riefer, 1990; Bayen et al., 1996).[1] In addition, Brainerd et al. (2012) proposed a new model that is able to overcome the reported shortcomings of the 1HTSM/2HTSM.

The manuscript is organized as follows: First, we will discuss the 1HTSM/2HTSM. This is followed by a characterization of source overdistribution and how it

* Corresponding author.
E-mail address: david.kellen@psychologie.uni-freiburg.de (D. Kellen).

[1] Throughout their manuscript, Brainerd et al. (2012) exclusively refer to the 1HTSM proposed by Batchelder and Riefer (1990). Still, the model equations presented include a distractor-detection parameter D_N belonging to the 2HTSM, which was later introduced by Bayen et al. (1996). Because of this confusion we attempt to refer to both models (1HTSM/2HTSM) when discussing Brainerd et al.'s claims. We also refer to the two models when discussing aspects that hold for both of them.

http://dx.doi.org/10.1016/j.jml.2014.07.001
0749-596X/

can be accounted for by the 1HTSM/2HTSM as well as by a new model proposed by Brainerd et al. (2012). We will show that Brainerd et al.'s dismissal of the 1HTSM/2HTSM as well as their characterization of memory processes made on the basis of their new model can be questioned on several grounds. Furthermore, we report results from a reanalysis of data from Brainerd et al. (2012), Brainerd, Wang, and Reyna (2013), and Yu and Bellezza (2000). Finally, we report two new experiments using an extended task in order to provide a better understanding of source overdistribution and its impact in memory modeling.

Modeling item and source memory data

The traditional method for evaluating source-memory consists of a study phase in which items with different contextual characteristics (i.e., different sources) such as words with different colors (e.g., red words and green words) are presented. Individuals subsequently engage in a test phase in which they have to distinguish between studied and non-studied words and identify the source of recognized items. There are typically two sources, A and B, and the response alternatives in this *source-memory task* are "Source A", "Source B", and "New".

The relationship between item and source-memory judgments is frequently modeled by means of measurement models designed to disentangle the contribution of different cognitive processes (Klauer & Kellen, 2010). One such model is the 2HTSM (Bayen et al., 1996), a model belonging to the Multinomial Processing Tree (MPT) model class (Batchelder & Riefer, 1999; Riefer & Batchelder, 1988). The 2HTSM assumes a finite set of discrete mental states that can be entered conditional on the occurrence of specific cognitive processes. The parameters in the model quantify the probability of each of these processes taking place (the parameter values are therefore bounded between 0 and 1). The 2HTSM is depicted in Fig. 1 for the case of two sources (A and B) in which the model assumes five mental states, from M_1 to M_5:

- M_1: An A item is remembered as previously studied and stemming from Source A.
- M_2: A B item is remembered as previously studied and stemming from Source B.
- M_3: An old item is remembered as previously studied, but memory for the source is absent.
- M_4: A new item is detected as new.
- M_5: An item presented at test is not remembered as previously studied nor is it detected as new.

The probability of each mental state being entered, given a particular type of item (A, B, or new), is determined by detection parameters that quantify the probability of specific memory processes successfully occurring: Parameters D_A and D_B correspond to the probability, respectively, of A and B items being remembered as previously studied (item memory). The memory-retrieval processes associated with D_A and D_B determine whether an item was previously studied or not, but not the source of these items. Parameter D_N captures the probability that a new item is actively rejected via so-called recall-to-reject processes (e.g., Rotello & Heit, 2000) or metacognitive strategies (e.g., Strack & Bless, 1994). In the restricted version of the 2HTSM, the 1HTSM, the possibility of an item being detected as new (M_4) is excluded (i.e., $D_N = 0$), so that responses to new items are completely governed by guessing processes, which will be described below.

Parameters d_A and d_B quantify the probability that the source of a studied item is remembered, conditional on item memory. The probability of the mental states described above being entered is a function of all of the above parameters: For example, for A items the probability of state M_1 corresponds to $D_A \times d_A$. From a broader perspective, the detection processes described by parameters D and d can be aligned with the familiarity and recollection processes postulated by dual-process models of memory (Klauer & Kellen, 2010; see also Malmberg, 2008).[2]

In states M_1, M_2, and M_4, the true status of the test item has been detected, which means that a correct response (responses "Source A", "Source B", and "New", respectively) is given with probability 1. In contrast, in M_3 and M_5 the true status of the test item is only partially detected or completely unknown, respectively. In these states, responses are guesses: State M_3 is mapped onto responses "Source A" and "Source B" with guessing probabilities γ_A and γ_B, with $\gamma_A + \gamma_B = 1$. In state M_5, responses "Source A", "Source B", and "New" are given with guessing probabilities β_A, β_B, and β_N, respectively, with $\beta_A + \beta_B + \beta_N = 1$. At this point, an important feature of the model should be emphasized: The detection processes modeled by parameters D_A, D_B, and D_N always provide accurate information, which means that the model does not permit judgment errors based on false information retrieved from memory. Consequently, the model attributes all observed errors to guessing processes, which are described by the γ and β parameters.[3] It is important to note that the exclusive assignment of errors to guessing processes does not mean that one is claiming that other types of errors (e.g., false recollection) do not exist. Rather that such types of errors do not constitute a major aspect of the particular data being characterized by the model (recognition memory judgments for non-related word lists associated to arbitrary contexts or sources). This is not expected to hold in the case of data coming from experimental paradigms where semantic/associative false memories are expected (for reviews, see Brainerd & Reyna, 2005; Gallo, 2006).

The model has been found to provide adequate fits to experimental data and has been experimentally validated:

[2] This comparison with dual-process models is done rather loosely and does not imply that the familiarity and recollection parameters in the dual-process model *exactly* correspond to item and source detection, respectively. As shown by Yu and Bellezza (2000, p. 1527), recollection and familiarity can be shown to be a complex function of item and source detection as well as guessing processes.

[3] This specification of the 1HTSM/2HTSM is actually a reparametrization of the one originally proposed by Batchelder and Riefer (1990), according to which there are two types of guessing judgments: (1) item-memory (old-new) judgments represented by parameter b, and (2) source-memory (Source A – Source B) guessing judgments represented by parameters a and g operating on mental states M_3 and M_5, respectively. The current model specification corresponds to Batchelder and Riefer's if $\gamma_A = a, \gamma_B = 1 - a$, $\beta_A = b \times g, \beta_B = b \times (1 - g)$, and $\beta_N = 1 - b$.

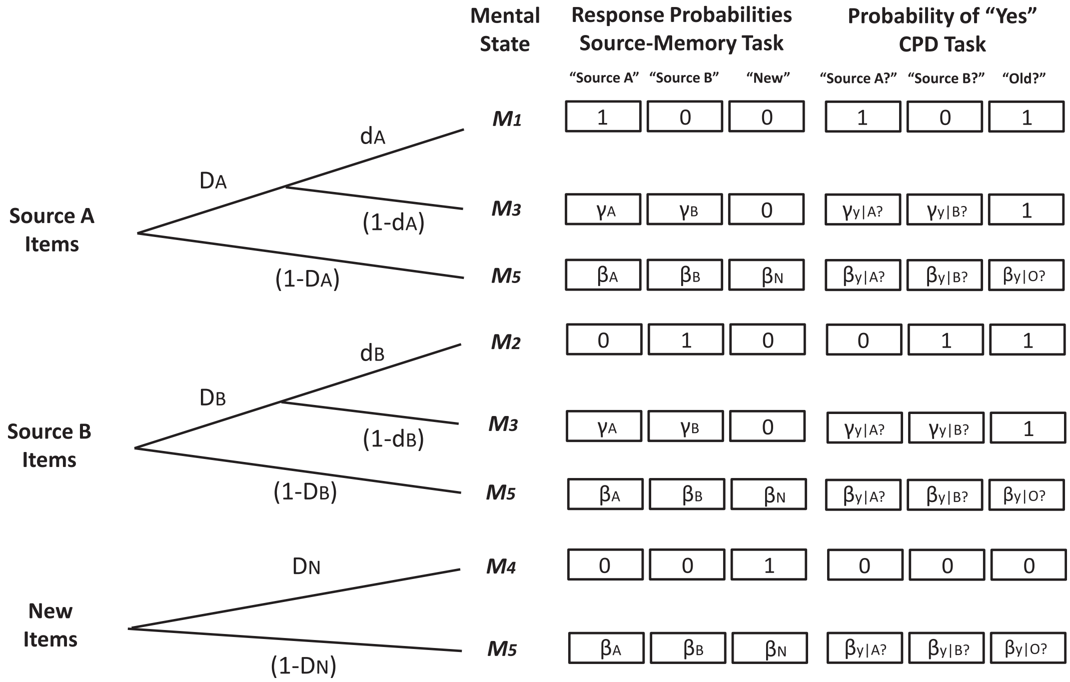

Fig. 1. Two-High Threshold Source-Memory Model (2HTSM) for a two-source experiment. Note that each depicted set of response probabilities associated to a mental state in the source-memory task sums to 1.

Several studies have shown that the different detection (D and d) and guessing parameters (β and γ) can be selectively influenced across experimental conditions (Bayen et al., 1996; Schütz & Bröder, 2011; Yu & Bellezza, 2000). Moreover, further studies have shown that the guessing tendencies captured by the β and γ parameters are affected in an expected manner by the probabilities associated to the different detection processes as well as the nature of the sources (Arnold, Bayen, Kuhlman, & Vaterrodt, 2013; Batchelder & Batchelder, 2008; Ehrenberg & Klauer, 2005; Klauer & Meiser, 2000; Meiser, 2003, 2005; Riefer, Hu, & Batchelder, 1994; Spaniol & Bayen, 2002). These results strongly suggest that the 2HTSM's parameters are successfully measuring the major cognitive processes underlying memory jugments in this paradigm.

The conjoint process-dissociation task, source overdistribution, and a new model for item and source memory

The source-memory task is only one possible way of obtaining item and source-memory judgments. One alternative is the *conjoint process-dissociation task* (henceforth CPD task; Brainerd & Reyna, 2008; Brainerd, Reyna, & Mojardin, 1999; Brainerd et al., 2012; Yu & Bellezza, 2000): In the CPD task, test items are presented along with one test probe that queries whether the item belongs to a specific source (e.g., “Source A?”) or that queries whether the item was previously studied (e.g., “Old?”). The participant is then required to respond “Yes” or “No” to the presented probe. Although less popular than the source-memory task, the CPD task is known to provide a comparable characterization of item and source-memory processes (see Yu & Bellezza (2000) for a demonstration of parameter equivalence across the source-memory and CPD tasks). The CPD task is an extension of Jacoby's (1991) original Process-Dissociation task, where items are tested in an *inclusion* (“Old?”) and an *exclusion* condition (“Source A?”).

Brainerd et al. (2012) argued that the proportion of “Yes” responses associated to the test probes “Source A?”, “Source B?”, and “Old?” can be seen as subjective estimates of source-membership probabilities, namely P(A), P(B), and $P(A \cup B)$. Following the laws of probability one can easily see that the (subjective) probability of the test item belonging to both sources is given by $P(A \cap B) = P(A) + P(B) - P(A \cup B)$. The probability of the conjunction — which Brainerd et al. term *source overdistribution* — is expected to be zero given that the sources are mutually exclusive, a fact that participants are made aware of in the instructions. Nevertheless, the experimental results reported by Brainerd et al. show that source overdistribution is consistently above zero. Brainerd et al. (2013) replicate these results. The phenomenon of source overdistribution is related to the phenomenon of episodic overdistribution in the recognition of semantically-related word lists (Brainerd & Reyna, 2008) as well as to a considerable body of results in the decision-making literature showing related biases in subjective-probability judgments (e.g., Rottenstreich & Tversky, 1997; Sloman, Rottenstreich, Wisniewski, & Hadjichristidis, 2004).

On the theoretical side, Brainerd et al. (2013) portrayed the phenomenon of overdistribution as a memory analog of the *quantum superposition principle*, according to which a physical system (e.g., an electron), prior to measurement or observation, exists simultaneously in all of its theoretically possible states, even when those states are mutually incompatible. This analogy is furthered by the classification of the CPD task as an experimental-psychological instantiation of the famous *double-slit experiment* from Physics (Feynman, Leighton, & Sands, 1965) where quantum superposition is observed: In the double-slit experiment, electrons are fired through two adjacent slits, with evidence (an interference pattern) indicating that each single electron is *simultaneously* passing through the two slits being found. Brainerd et al. (2013) argue that superposition in memory manifests itself via a deterministic acceptance of the test item as member of or satisfying any test probe irrespective of their mutual incompatibility, similar to the way single electrons simultaneously pass through the slits. This apparent similarity between memory overdistribution and quantum superposition has in fact been used as a motivation for quantum-probability modeling in Psychology (Brainerd et al., 2013; Busemeyer & Bruza, 2012; Busemeyer & Trueblood, 2010). In fact, proponents of quantum-probability models argue that the phenomenon of source overdistribution is "puzzling" from a classical probability perspective (Wang, Busemeyer, Atmanspacher, & Pothos, 2013, p. 681).

According to Brainerd, Holliday, Nakamura, and Reyna (in press), when an item is remembered but the source is not retrieved there is a combination of confidence and uncertainty in the sense that the individual is sure that the item was previously studied but he/she is not able to remember its context. In fact, the evidence available is consistent with any of the mutually exclusive contexts/sources. This consistency leads participants to attribute the test item to the source described in the test probe. The same applies to the case of semantically-related word lists (Brainerd & Reyna, 2008), where the phenomenon of overdistribution is associated with the retrieval of gist-memory traces that produce a generic phenomenology — semantic similarity — which is consistent with different episodic states (studied items versus related distractors), thus leading to a general acceptance of test probes.

In order to account for the observed overdistribution in the CPD task, Brainerd et al. (2012) proposed the Conjoint Process-Dissociation Model (CPDM). The discrete mental states assumed by the CPDM differ from the ones assumed by the 1HTSM/2HTSM. The CPDM is depicted in Fig. 2 for the case of two sources. For two sources, the model assumes six mental states, labeled W_1 to W_6:

- W_1: An A item is correctly recollected as coming from Source A.
- W_2: An A item is erroneously recollected as coming from Source B.
- W_3: A B item is correctly recollected as coming from Source B.
- W_4: A B item is erroneously recollected as coming from Source A.
- W_5: An old item is remembered as being previously studied, and attributed to *whatever* source is being probed.
- W_6: For an item presented at test, no information is available.

Parameters R_A and R_B represent the probability of the source of A and B items being correctly remembered,

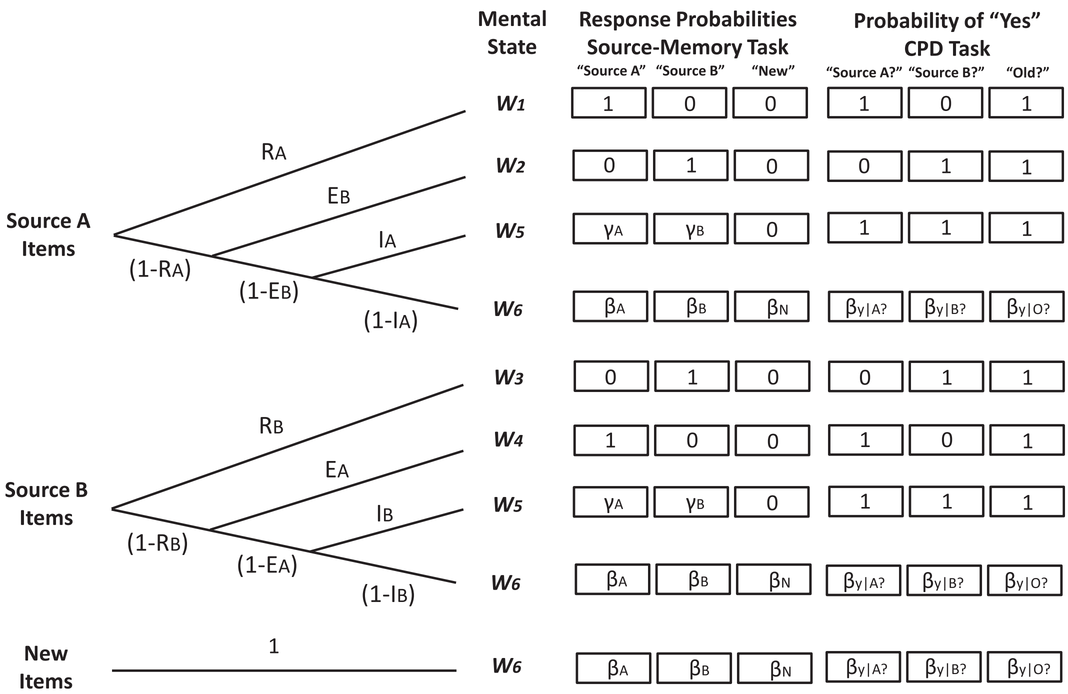

Fig. 2. Conjoint Process-Dissociation Model (CPDM) for a two-source experiment. Note that each depicted set of response probabilities associated to a mental state in the source-memory task sums to 1.

respectively. In contrast, parameters E_B and E_A correspond to the (conditional) probability of A and B items being incorrectly remembered as coming from Source B and Source A, respectively. Parameters I_A and I_B represent the (conditional) probability that only item memory is available, in which case individuals will respond "Yes" to the item whichever the probe, *invariably* attributing it to the source referred to in the test probe (Brainerd et al., 2012, p. 415). Thus, when only item memory is available, the item would be attributed to both sources, which is a logically inconsistent response pattern. Finally, parameters $\beta_{y|A?}$ $\beta_{y|B?}$, and $\beta_{y|O?}$ describe the guessing tendencies to respond "Yes" for probes "Source A?", "Source B?", and "Old?" respectively.

Certain aspects of the CPDM should be emphasized: First, the model assumes that errors can (also) occur due to false recollection (e.g., Dodson, Bawa, & Slotnick, 2007), which contrasts with the 2HTSM's assumption that errors are solely due to guessing. Second, the item-memory state (W_5) postulated by the model assumes that the test item is *invariably* accepted, *independently* of the test probe (see also Brainerd & Reyna, 2008). Thus, the model excludes the possibility that individuals remember a given test item but are not certain about its source(s), and then choose to express this uncertainty by responding "No" rather than "Yes" at least sometimes. This assumption of CPDM is not only in contrast with the 2HTSM but also with bivariate SDT and dual-process approaches to item and source-memory (Hautus et al., 2008; Klauer & Kellen, 2010; Onyper et al., 2010) which do not assume such a deterministic process ("Yes" responses are *probabilistic* and depend on additional aspects such as subjects' *response criteria*, for example). Furthermore, this assumption is also at odds with research pointing to the existence of different source-guessing strategies when only item memory is available (Batchelder & Batchelder, 2008; Riefer et al., 1994). Brainerd et al. (2012) argue that when only item information is available, retrieval processes attribute the test item to too many contexts (i.e., a case of memory superposition), even when these memory attributions represent logical contradictions.

According to the CPDM for a two-source CPD task, the probability of response "Yes" for the different probes (for the case of A items) is as follows for A and new items (for the sake of brevity, we leave out the analogous equations for B items):

$$\begin{aligned}
P(\text{Yes}|\text{A?}, \text{A}) &= R_A + (1-R_A)(1-E_B)I_A \\
&\quad + (1-R_A)(1-E_B)(1-I_A)\beta_{y|A?}, \\
P(\text{Yes}|\text{B?}, \text{A}) &= (1-R_A)E_B + (1-R_A)(1-E_B)I_A \\
&\quad + (1-R_A)(1-E_B)(1-I_A)\beta_{y|B?}, \\
P(\text{Yes}|\text{Old?}, \text{A}) &= 1-(1-R_A)(1-E_B)(1-I_A)(1-\beta_{y|O?}), \\
P(\text{Yes}|\text{A?}, \text{New}) &= \beta_{y|A?}, \\
P(\text{Yes}|\text{B?}, \text{New}) &= \beta_{y|B?}, \\
P(\text{Yes}|\text{Old?}, \text{New}) &= \beta_{y|O?}.
\end{aligned}$$

Model equations for both two and three-source CPD tasks can be found in the Supplemental Material. Now, let us consider the case of overdistribution for A and New items, whose expressions can be obtained with basic algebraic manipulations:

$$\begin{aligned}
P(\text{A} \cap \text{B}|\text{A}) &= (1-R_A)(1-E_B)I_A + (1-R_A)(1-E_B)(1-I_A) \\
&\quad \times (\beta_{y|A?} + \beta_{y|B?} - \beta_{y|O?}),
\end{aligned} \tag{1}$$

$$P(\text{A} \cap \text{B}|\text{New}) = \beta_{y|A?} + \beta_{y|B?} - \beta_{y|O?}. \tag{2}$$

The first term in Eq. (1) of item memory in the absence of recollection. The second term shows that overdistribution can also be caused by differences in guessing tendencies across test probes, particularly when $\beta_{y|A?} + \beta_{y|B?} > \beta_{y|O?}$.[4] Eq. (2) shows that this inequality between the guessing parameters will also lead to the observation of source overdistribution in the case of new items. Note as one example that this inequality between the guessing parameters is expected to hold when individuals do not perfectly adjust (or do not adjust at all) their guessing tendencies to the proportion of items in the test list that match the test probe (e.g., Cox & Dobbins, 2011; Thomas & Legge, 1970). For example, if all three β parameters are of equal size, overdistribution is predicted as a consequence of guessing processes. Still, note that the CPDM can also predict negative overdistribution if item memory is low and $\beta_{y|A?} + \beta_{y|B?} < \beta_{y|O?}$. Finally, note that source overdistribution is expected to be greater when memory is poor (especially source memory). However, given that source overdistribution is the outcome of several different processes, the possibility of interactions among the latter compromises the establishment of clear-cut predictions.

One of the predictions made by the CPDM (and confirmed empirically by Brainerd et al. (2012)) is that source overdistribution increases as the number of sources used increases. For example, if three sources (A, B, and C) are used, the CPDM predicts

$$\begin{aligned}
P(\text{A} \cap \text{B} \cap \text{C}|\text{A}) &= 2 \times (1-R_A)(1-E_B)(1-E_C)I_A + (1-R_A) \\
&\quad \times (1-E_B)(1-E_C)(1-I_A) \times (\beta_{y|A?} + \beta_{y|B?} \\
&\quad + \beta_{y|C?} - \beta_{y|O?}),
\end{aligned}$$

where the first term, corresponding to the contribution of item memory to overdistribution, is doubled. From this it is easy to see how overdistribution is expected to be larger in the case of three sources. First, assume that all parameters shared by the two overdistribution equations (e.g., R_A) have the same values. Now, if false-recollection parameter E_C is smaller than .50 (something which is highly expected; see Brainerd, Wright, Reyna, & Mojardin, 2001), then overdistribution is expected to be larger in the three-source case than in the two-source case.

[4] Note that when the task format enforces $\beta_{y|A?} + \beta_{y|B?} + \beta_{y|O?} = 1$ and $\gamma_{y|A?} + \gamma_{y|B?} = 1$ as when participants have to respond with one of three response options, (1) that the item was studied in and belongs to Source A, (2) that the item was studied in and belongs to Source B, (3) that the item is new (see Batchelder & Riefer, 1990; see also Footnote 3), then the predicted overdistribution due to guessing processes is 0. Batchelder and Riefer's parametrization is a special case of the parametrization used here, a parametrization which is the most general and allows for the 1HTSM/2HTSM to be tested on the basis of its core assumptions (the mental states M).

Revisiting Brainerd et al.'s (2012) model analysis

A new model is usually shown to outperform other models already established in the literature before being accepted for further scientific scrutiny. In the case of source-memory modeling, an obvious point of comparison would be either the 1HTSM or the 2HTSM. Brainerd et al. (2012) reported a comparison between the CPDM and the 1HTSM/2HTSM (see Footnote 1) and correctly pointed out that the latter model attributes source overdistribution entirely to guessing processes. Furthermore, Brainerd et al. (2012) fitted both models to the experimental data reported and found that the 1HTSM/2HTSM provided a much poorer account, producing gross misfits that were rejected under null-hypothesis testing. According to Brainerd et al., these results indicate that the 1HTSM/2HTSM, but not the CPDM, mischaracterizes the underlying cognitive processes, which are seemingly better captured by the CPDM. As shown next, the comparison reported by Brainerd et al. (2012) has important limitations that question the conclusiveness of the reported model comparisons as well as the related claims regarding the dependence of cognitive processes: First, we will show that Brainerd et al. (2012) considered an incorrect specification of the 1HTSM/2HTSM. Second, we will demonstrate that, contrary to Brainerd et al.'s claims, the CPDM makes no assumptions regarding the relationship between item memory and source recollection. Finally, we reanalyze previously-published CPD-task data (Brainerd et al., 2012, 2013; Yu & Bellezza, 2000) with the 1HTSM and CPDM and show that both models are very similar in their ability to account for the data, contradicting the notion that the 1HTSM/2HTSM is unable to account for CPD-task data.

Incorrect specification of the 1HTSM/2HTSM

Let us begin with the misspecification of the 1HTSM/2HTSM. Brainerd et al. (2012) fitted the models to each source separately. In the case of two-source CPD data, responses given to items from one list (e.g., A items) and to new items together provide six degrees of freedom, a limited number that imposes the need to fit restricted versions of the 1HTSM/2HTSM in order to have models that are both *testable* (can be rejected by the data) and *identifiable* (the account provided has a unique set of parameters; for a discussion, see Bamber & Van Santen, 2000).

Brainerd et al. (2012) specify the equations of the 1HTSM/2HTSM for the 2-source CPD task as follows:

$$\begin{aligned}
&\mathrm{P(Yes|A?,A)} = D_A d_A + D_A(1-d_A)\gamma_y,\\
&\mathrm{P(Yes|B?,A)} = D_A(1-d_A)\gamma_y,\\
&\mathrm{P(Yes|Old?,A)} = D_A,\\
&\mathrm{P(Yes|A?,New)} = (1-D_N)\beta_y,\\
&\mathrm{P(Yes|B?,New)} = (1-D_N)\beta_y,\\
&\mathrm{P(Yes|Old?,New)} = (1-D_N)\beta_y.
\end{aligned}$$

Given the equations above, the following expression for source overdistribution in A items results:

$$\mathrm{P(A\cap B|A)} = D_A(1-d_A)(2\gamma_y-1), \tag{3}$$

$$\mathrm{P(A\cap B|N)} = (1-D_N)\beta_y. \tag{4}$$

According to Eq. (3), source overdistribution is entirely produced by a specific guessing process, and is expected to be above 0 when $\gamma_y > .50$. Eq. (4) new items, source overdistribution simply corresponds to the predicted false-alarm rate.

At this point it is important to note that Brainerd et al.'s (2012) specification of the 1HTSM/2HTSM leading to Eqs. (3) and (4) considerably departs from traditional specifications in several ways: First, it precludes any form of guessing when no memory is available, which occurs, for example, for A items with probability $1-D_A$. Instead, it is tacitly assumed for studied items that response "No" is *deterministically* given when no memory is available (state M_5). This assumption is problematic given that it reduces the number of ways that source overdistribution can come about in the 2HTSM, compromising the validity of any comparison. Also of note is the fact that although M_5 invariably leads to "No" for the case of old items, a guessing process is assumed for the case of new items. One of the core assumptions of the 2HTSM is, however, that individuals behave similarly towards items in state M_5, whether they are old or new. That is, the impact of the item type (e.g., studied or non-studied) on responses is expected to be completely mediated by the mental state that is entered, a principle that Brainerd et al.'s specification violates without any sort of justification.

The specification of the 1HTSM/2HTSM for new items has additional problems: The guessing process that occurs in the absence of distractor rejection is assumed to be invariant across test probes, with the probability of response "Yes" being $(1-D_N)\beta_y$ for all test probes. In contrast, no such restriction is present in the CPDM, which reasonably permits different β parameters for the different test probes (see Brainerd et al., 2012, Appendix). The same problems can be found in the implementation of the 1HTSM/2HTSM for the three-source CPD task.

Overall, the evaluation of the 1HTSM/2HTSM hinges upon an implementation that does not have some of the defining properties of the model, and therefore has limited diagnostic value regarding the ability of the 1HTSM/2HTSM to account for data in general, let alone source overdistribution.

A specification of the 1HTSM/2HTSM for the CPD task that is in line with the original model for the source-memory task (a) uses the same mental states M_1 to M_5 that characterize memory states in the source-memory task and (b) adjusts the state-response mapping to suit the demands of the CPD task (state-response mappings are shown in the rectangular boxes of Fig. 1):

$$\begin{aligned}
&\mathrm{P(Yes|A?,A)} = D_A d_A + D_A(1-d_A)\gamma_{y|A?} + (1-D_A)\beta_{y|A?},\\
&\mathrm{P(Yes|B?,A)} = D_A(1-d_A)\gamma_{y|B?} + (1-D_A)\beta_{y|B?},\\
&\mathrm{P(Yes|Old?,A)} = D_A + (1-D_A)\beta_{y|O?},\\
&\mathrm{P(Yes|A?,New)} = (1-D_N)\beta_{y|A?},\\
&\mathrm{P(Yes|B?,New)} = (1-D_N)\beta_{y|B?},\\
&\mathrm{P(Yes|Old?,New)} = (1-D_N)\beta_{y|O?}.
\end{aligned}$$

This specification in turn leads to

$$P(A \cap B|A) = D_A(1 - d_A)(\gamma_{y|A?} + \gamma_{y|B?} - 1) + (1 - D_A) \times (\beta_{y|A?} + \beta_{y|B?} - \beta_{y|O?}), \quad (5)$$

$$P(A \cap B|N) = (1 - D_N)(\beta_{y|A?} + \beta_{y|B?} - \beta_{y|O?}). \quad (6)$$

Eqs. (5) and (6) show that the 1HTSM/2HTSM predicts source overdistribution via two different kinds of guessing processes, when $\gamma_{y|A?} + \gamma_{y|B?} > 1$ and/or $\beta_{y|A?} + \beta_{y|B?} > \beta_{y|O?}$. As previously stated when discussing the CPDM's guessing (β) parameters, inequalities of this kind are expected to hold if participants do not adjust their guessing rates to match the base-rate of the item being probed. Given that individuals are known to be quite resistant to adjust their response tendencies (e.g., Cox & Dobbins, 2011) and prone to under-adjustment or response conservatism (e.g., Thomas & Legge, 1970), the occurrence of a guessing-based overdistribution phenomenon is expected. Previous work on the Process Dissociation task (Jacoby, 1991) has focused on false-alarm differences (which are assumed to reflect guessing differences) across test probes, with some studies showing no differences in false alarms across test probes (e.g., Jacoby, 1991) while others showing (small, but consistent) changes (e.g., Yu & Bellezza, 2000). Eqs. (5) and (6) show that these observed false-alarm rates would lead to the observation of overdistribution. Note that the 1HTSM/2HTSM can also predict negative overdistribution if $\gamma_{y|A?} + \gamma_{y|B?} < 1$ and/or $\beta_{y|A?} + \beta_{y|B?} < \beta_{y|O?}$. Like the CPDM, the 1HTSM/2HTSM's characterization of overdistribution leads to the prediction that the latter should be greater when memory is poor. However, it is possible that changes in other parameters (e.g., guessing) compromise the observation of such relationship.

Besides the ability to account for source overdistribution in a two-source CPD task, the 1HTSM/2HTSM is also able to accomodate its increase in the case of three sources:

$$P(A \cap B \cap C|A) = D_A(1 - d_A)(\gamma_{y|A?} + \gamma_{y|B?} + \gamma_{y|C?} - 1) + (1 - D_A)(\beta_{y|A?} + \beta_{y|B?} + \beta_{y|C?} - \beta_{y|O?}).$$

Overdistribution is expected when $\beta_{y|A?} + \beta_{y|B?} + \beta_{y|C?} > \beta_{y|O?}$ and/or $\gamma_{y|A?} + \gamma_{y|B?} + \gamma_{y|C?} > 1$. As for the CPDM, the sum of three guessing parameters (e.g., $\gamma_{y|A?}, \gamma_{y|B?}$, and $\gamma_{y|C?}$) should produce a larger overdistribution than the sum of only two (e.g., $\gamma_{y|A?}$ and $\gamma_{y|B?}$).

Dependency of mental processes

Brainerd et al. (2012) argue that the CPDM imposes a relationship over retrieval states — that recollection states are *independent* from item memory and thus can occur in the absence of the latter. Brainerd et al. (2012) refitted their datasets with a new version of CPDM in which source recollection depended on the successful occurrence of item memory, and found that this model grossly misfitted the data. These results led Brainerd et al. to claim that recollection can occur in the absence of item memory, contradicting the common modeling assumption that source memory is dependent on item memory (Hautus et al., 2008; Klauer & Kellen, 2010). The goal of this section is to show that Brainerd et al.'s conclusions are premature as the CPDM is silent regarding the dependency of item and source memory. Particularly, we will demonstrate that an alternative model that only assumes recollection in the presence of item memory is formally equivalent to the CPDM.

Before focusing on the CPDM, a brief discussion on how previous modeling accounts like the 1HTSM/2HTSM capture the empirical relationship between item and source memory is in order: The possibility of source memory in the absence of item recognition is precluded by models such as the 1HTSM/2HTSM, but this assumption is actually immaterial in the context of traditional source-memory tasks (e.g., Batchelder & Riefer, 1990) where source judgments are only required for recognized items. Any adjustment of the 1HTSM/2HTSM would be inconsequential as the outcome of any tree branching representing source memory in the absence of item memory would not lead to an observable response: For example, branching $(1 - D_A) \times (1 - \beta) \times \hat{d}_A$ with $\hat{d}_A$ denoting the probability of source memory for non-recognized A items, would never be observed given that participants do not have the possibility to assign an item to a source without stating explicitly or implicitly that they remember it as old in the traditional source discrimination task. Such kind of responses only recently became admissible with the modeling of confidence-rating judgments (Hautus et al., 2008; Klauer & Kellen, 2010) where individuals provided source judgments independently of their item-memory judgment. However, no evidence for the presence of source-memory for non-recognized items was actually found in those studies. This lack of evidence is particularly interesting given that it contrasts with focused studies (e.g., Ceci, Fitneva, & Williams, 2010; Starns, Hicks, Brown, & Martin, 2008) showing evidence for above-chance source-memory accuracy for non-recognized items (although this accuracy is still lower than in the case of recognized items). For example, Starns et al. argue that their results support a two-dimensional signal detection theory account of item and source memory where item and source memory are positively correlated (DeCarlo, 2003), but the modeling work of Hautus et al. (2008) shows that a post hoc adjustment imposing chance-level source judgments for non-recognized items is necessary in order for the signal detection model to give a reasonable account of the data. This discrepancy suggests that source memory in the absence of item memory does not contribute substantially to the experimental paradigms commonly used in source-memory modeling endeavors (see Klauer & Kellen, 2010) and only needs to be taken into account in focused tests, if at all. In any event, the pronounced differences observed in source-memory accuracy between recognized and non-recognized items across the different studies already indicate that item and source memory processes are unlikely to be stochastically independent (otherwise, source-memory accuracy should be exactly the same for recognized and non-recognized items).

We now turn to the tree structure of the CPDM. Brainerd et al.'s (2012) claim that the implied dependency/independence of processes in the CPDM is meaningful in the current setting is inaccurate. As it turns out, the specification of processes in the CPDM can be shown to be completely *silent* regarding the actual dependence/independence of mental processes. As an example, let us consider a new model for CPD-task data, entitled CPDM*, in which item memory not only precedes but is also a requirement for the occurrence of recollection:

$$\begin{aligned}
&\mathrm{P(Yes|A?,A)} = I_A^* R_A^* + I_A^*(1 - R_A^*)(1 - E_B^*) + (1 - I_A^*)\beta^*_{y|A?},\\
&\mathrm{P(Yes|B?,A)} = I_A^*(1 - R_A^*) + (1 - I_A^*)\beta^*_{y|B?},\\
&\mathrm{P(Yes|Old?,A)} = I_A^* + (1 - I_A^*)\beta^*_{y|O?}
\end{aligned}$$

The meaning of the parameters in CPDM* is slightly different from CPDM: For example, while I_A^* represents the probability of item memory being available, I_A is the probability of item memory *conditional* on the absence of both true and false recollection. On the other hand, while R_A represents the probability of recollection, R_A^* represents the probability of recollection *conditional* on the presence of item memory. Despite the slight differences in the interpretation of the parameters, both the CPDM and CPDM* are just different ways of describing the same joint events. With some basic algebraic manipulation it is easy to show that the change in the order of the parameters only amounts to a reparametrization of the same multinomial distributions, which means that the parameters of one model can be set as a function of the other without any loss of generality:

$$\begin{aligned}
&I_A^* = R_A + (1 - R_A)E_B + (1 - R_A)(1 - E_B)I_A,\\
&R_A^* = \frac{R_A}{R_A + (1 - R_A)E_B + (1 - R_A)(1 - E_B)I_A},\\
&E_B^* = \frac{(1 - R_A)E_B}{(1 - R_A)E_B + (1 - R_A)(1 - E_B)I_A},
\end{aligned}$$

or conversely

$$\begin{aligned}
&I_A = \frac{(1 - E_B^*)(1 - R_A^*)I_A^*}{(1 - I_A^*) + I_A^*(1 - R_A^*)(1 - E_B^*)},\\
&R_A = R_A^* I_A^*,\\
&E_B = \frac{E_B^*(1 - R_A^*)I_A^*}{1 - R_A^* I_A^*}.
\end{aligned}$$

CPDM and CPDM* are simple reparametrizations of each other, so that their goodness of fit to actual data is the same despite the different tree structure in the model specification. The reason for the misfits reported by Brainerd et al. (2012) for the rearranged model stem from the fact that the *E* parameters were omitted in their CPDM* equations (see Brainerd et al., Equations A15-A20). An evaluation of process dependence requires that the same processes are present in the two models. The demonstrated CPDM/CPDM* equivalence is especially important because it shows that the order in which the processes occur in the model do not lead to testable differences. This means that the CPDM's ability to fit data from the CPD task (or its performance relative to the 1HTSM/2HTSM) *has no bearing whatsoever* on the presumed relationship between item and source memory.

A similar point was made by Buchner, Erdfelder, and Vaterrodt-Plünnecke (1995) for the case of Jacoby's (1991) Process Dissociation model (a conceptual predecessor of the CPDM; Brainerd et al., 1999, 2012), where they showed that the stochastic relationship between the recollection and familiarity processes was not identifiable (see also Buchner & Erdfelder, 1996) as the characterization of observed responses provided by the model equations were simultaneously consistent with process independence, mutually exclusivity, and redundancy (Jones, 1987). This non-identifiability issue in the Process Dissociation model emerges from the fact that one can only estimate familiarity in the absence of recollection whereas the probability of familiarity given recollection is not observable (at least in traditional Process Dissociation designs). This non-identifiability issue should not be confused with the discussion regarding whether or not recollection and familiarity estimates are correlated when aggregating responses across participants or items (e.g., Curran & Hintzman, 1995, 1997; Jacoby & Shrout, 1997). Finally, note that these non-identifiability issues are not always present in models of this kind, as there are models (together with their associated experimental designs) where different process orderings and dependencies can be distinguished and tested (see Schweickert & Chen, 2008; Schweickert & Xi, 2011).

Can we reinterpret the CPDM parameters in terms of the 1HTSM/2HTSM?

In the previous section we focused on the dependency of the processes postulated by the CPDM. We will now discuss a related issue, the possibility of reinterpreting the CPDM false recollection (*E*) and item-memory/superposition (*I*) parameters in terms of the guessing processes postulated by the 1HTSM/2HTSM.[5]

Let us begin with false-recollection parameter *E*: This parameter *cannot* be reinterpreted as a guessing parameter like γ because it can assume different values for the different sources (e.g., $E_A \neq E_B$). A source-guessing process implies that it takes on the same values across sources, given that source information is expected to be absent (after all, it is this absence of information that creates the need to guess).

We now turn to the item-memory/superposition parameter *I*. This parameter establishes the (conditional) probability of the test item being in a memory state in which a "Yes" response is invariably produced. In contrast with the 1HTSM/2HTSM, item memory is not mapped onto observed responses via a guessing process. Instead, acceptance of the test probe results from the retrieved evidence from item memory that is postulated to be consistent with all the test probes simultaneously. This absence of a guessing process in item memory reflects a theoretical position that distinguishes the CPDM from the 1HTSM/2HTSM:

> At the level of retrieval processes, the difference between the [CPDM] and [1HTSM/2HTSM] accounts of overdistribution is just this. On the one hand, [CPDM] localizes the cause of overdistribution within item memory, positing that overdistribution will occur for cues for which subjects have item memory when subjects cannot recollect the cues' sources. On the other

[5] We thank Charles J. Brainerd for suggesting this comparison.

> hand, the [1HTSM/2HTSM] localizes overdistribution within a guessing process, positing that for cues for which subjects do not have item memory, overdistribution occurs when subjects tend to guess the sources that are stipulated in exclusion and mirror exclusion probes at high rates (Brainerd et al., 2012, p. 416)

Still, one could in principle introduce a guessing process in the case of item memory into the CPDM. For the two-source CPD task, the CPDM formula for overdistribution would then be

$$P(A\cap B|A) = (1-R_A)(1-E_B)I_A(\gamma_{y|A?}+\gamma_{y|B?}-1) + (1-R_A)(1-E_B)(1-I_A)(\beta_{y|A?}+\beta_{y|B?}-\beta_{y|O?}). \tag{7}$$

Eq. (7) is equivalent to Eq. (1) when $\gamma_{y|A?}=\gamma_{y|B?}=1$. Thus, the CPDM for two sources can be seen as arising from a restriction on the guessing processes that invariably produces "Yes" responses. In general terms, the guessing-extended CPDM is consistent with Brainerd et al.'s (2012) CPDM when all γ parameters are fixed to $\frac{2}{S}$, where $S \geqslant 2$ is the number of sources. Note that when there are more than two sources, this guessing process no longer produces a deterministic guessing acceptance of the test probe (however, it is still imposes a fixed guessing rate).

This comparison highlights the similarities between the two models and the possibility of integrating them both within an encompassing model. Let us take the CPDM* and extend it with the item-memory guessing processes discussed above. The resulting equations for the two-source CPD task are

$$\begin{aligned}
P(\text{Yes}|A?,A) &= I_A^*R_A^* + I_A^*(1-R_A^*)(1-E_B^*)\gamma^*_{y|A?} + (1-I_A^*)\beta^*_{y|A?},\\
P(\text{Yes}|B?,A) &= I_A^*(1-R_A^*)E_B^* + I_A^*(1-R_A^*)(1-E_B^*)\gamma^*_{y|B?}\\
&\quad + (1-I_A^*)\beta^*_{y|B?},\\
P(\text{Yes}|\text{Old}?,A) &= I_A^* + (1-I_A^*)\beta^*_{y|O?}.
\end{aligned}$$

By setting $I_A^*=D_A, R_A^*=d_A$, and $E_B^*=0$, the equations become equivalent to the 1HTSM/2HTSM equations. In contrast, they are equivalent to the CPDM* equations when the γ^* parameters are fixed to 1. Thus the encompassing model can be seen as an extended 1HTSM/2HTSM where false recollection is possible. Furthermore, the encompassing model and its restricted cases assert and clarify the main diverging features of the two models: (1) the possibility of false recollection, and (2) the deterministic (or fixed) acceptance rates for item memory in the CPDM, versus the probabilistic guessing processes in the 1HTSM/2HTSM. Unfortunately, due to its complexity the encompassing model cannot be estimated in the experimental settings discussed in this manuscript and in the literature at large.

Comparing between the 1HTSM and the CPDM empirically: a reanalysis of CPD-task data

In this section we conduct a reanalysis of CPD-task data using correctly-specified versions of the 1HTSM/2HTSM. One additional problem in Brainerd et al.'s (2012) analysis is that the goodness-of-fit results for each source/list were summed across all stimulus conditions. This procedure is problematic because the responses to a common set of new items are "recycled" when modeling the responses to each studied list, violating the assumption of independent observations that is required for the summed fits to be meaningful (Bishop, Fienberg, & Holland, 1975).

In order to sidestep this assumption violation we will instead fit models to data from all sources and new items simultaneously, an approach that has been frequently advocated in the source-memory literature (e.g., Bröder & Meiser, 2007; DeCarlo, 2003; Hautus et al., 2008; Klauer & Kellen, 2010; Slotnick & Dodson, 2005). The two and three-source aggregate datasets from Brainerd et al. (2012, 2013), and Yu and Bellezza (2000) will be analyzed with the 1HTSM and CPDM.[6] The 1HTSM is used because this model is identifiable and testable with two and three-source CPD-task data.

The evaluation of relative model performance will be based on the *Normalized Maximum Likelihood* index (NML; Myung, Navarro, & Pitt, 2006) in order to take possible differences in model flexibility into account. NML is a model selection index emerging from the Minimum Description Length framework (MDL; Grünwald, 2007). According to MDL, data can be seen as a code whose length can be compressed by a model (itself a code with a particular length) according to the regularities present in the data. The (logarithm of the) NML index for an arbitrary model is given by:

$$\text{NML} = -\log f(x|\hat{\theta}(x)) + \log \sum_y f(y|\hat{\theta}(y)).$$

The first term corresponds to the model's maximum log-likelihood for observed data x and quantifies goodness of fit. The second term is a penalty factor that is the sum of the maximum log-likelihoods of all possible data patterns y that could in principle be observed in such experiment. In other words, the first term quantifies a model's goodness of fit (in terms of the maximum likelihood) and the second term penalizes the model for its ability to account for any data that might be observed (again in terms of maximum likelihoods). The larger the flexibility the larger the penalty. NML has been shown to outperform several alternative model-selection methods, and even approximate optimal model-recovery rates (e.g., Kellen, Klauer, & Bröder, 2013; Klauer & Kellen, 2010). The computation of NML represents a formidable challenge as it becomes intractable even for moderate sample sizes (alternative MDL-based measures such as the Fisher Information Approximation could be used in those cases as they are readily available for MPT model class; see Wu, Myung, & Batchelder, 2010a, 2010b; but see also Navarro, 2004). Klauer and Kellen (2011) developed methods for estimating NML in binomial and joint-binomial data that will be used in the present case.

[6] A couple of minor data-entry errors in the new-item trials were found in the two-source CPT data from Brainerd et al. (2012). The correction of these frequencies was obvious due to the expected number of trials and the fact that the frequencies for the distractor trials appeared twice in the data files (once with each source).

Both the 1HTSM and the CPDM are saturated in the two-source CPD task (nine parameters for nine degrees of freedom) but not in the three-source task (both have thirteen parameters for sixteen degrees of freedom). Although the two models are saturated in the two-source task they are still testable because they nevertheless impose inequality restrictions on the data (i.e. there are data patterns that they are unable to account for; for a similar case, see Kellen & Klauer, 2011). In such cases, the sampling distribution of the goodness-of-fit statistic G^2 follows a mixture of χ^2 distributions with the most conservative distribution being $\frac{1}{2}\chi^2_{df=1}$, which has a critical value ($p = .05$) of 2.71 (e.g., Self & Liang, 1987; for an example, see Regenwetter & Davis-Stober, 2012).

Model performance is summarized in Table 1: Both the 1HTSM and the CPDM are frequently rejected ($p < .05$), indicating that both models are failing to account for many of the observed data (see Note to Table 1). Both models have often similar goodness of fit values, which suggests that the shortcomings of both models do not result from their different assumptions. Regarding the predicted levels of source overdistribution, the results shown in Table 1 indicate that both models produced very similar predictions, which were very often close to the observed overdistribution values. In fact, the absolute agreement (measured by the Concordance Correlation Coeficient (CCC); see Barchard, 2012) between the two models' predicted overdistribution values was virtually perfect (CCC = 0.996). The observed and predicted overdistribution values are plotted in Fig. 3. Parameter estimates and their respective confidence intervals are reported in the Supplemental Material.

In terms of non-nested model selection, the computed NML values provide mixed results, as the 1HTSM is preferred in the two-source data and the CPDM is preferred in the three-source data. Overall, there is a preference for the CPDM, but it is not systematic according to a (permutation-based) Wilcoxon test (summed ΔNML = 10.60, $Z = 0.39, p = .73$; see Hothorn, Hornik, van de Wiel, & Zeileis, 2008). According to the NML metric, the 1HTSM is less flexible than the CPDM in the two-source CPD task (average penalty difference is -0.81), but the opposite holds in the three-source task (average penalty difference is 0.54). This difference indicates that model flexibility is dependent on the task used (see Kellen & Klauer, 2011; Kellen et al., 2013). Still, one aspect that needs to be taken into account is that flexibility is of secondary importance when none of the models provides a reasonable account of the data to begin with. If one only considers the cases where at least one of the models is not rejected by the data,

Table 1
Model performance.

Data	1HTSM			CPDM			Observed overdistribution
	G^2	NML	Predicted overdistribution	G^2	NML	Predicted overdistribution	
Two-source CPD task							
B1-Exp.1 HF-HC	**3.41**	24.28	[0.33, 0.25, 0.25]	0.10	23.47	[0.37, 0.17, 0.25]	[0.38, 0.17, 0.25]
B1-Exp.1 HF-LC	1.56	23.34	[0.47, 0.39, 0.50]	0.00	23.41	[0.50, 0.34, 0.50]	[0.50, 0.34, 0.50]
B1-Exp.1 LF-HC	0.00	22.56	[0.37, 0.31, 0.13]	0.00	23.41	[0.37, 0.31, 0.13]	[0.37, 0.31, 0.13]
B1-Exp.1 LF-LC	**17.43**	31.22	[0.54, 0.51, 0.33]	**17.43**	32.08	[0.54, 0.51, 0.33]	[0.54, 0.69, 0.29]
B1-Exp.2 HF-HC	0.00	22.48	[0.23, 0.22, 0.26]	0.00	23.34	[0.23, 0.22, 0.26]	[0.23, 0.22, 0.26]
B1-Exp.2 HF-LC	1.86	23.41	[0.29, 0.27, 0.30]	1.86	24.27	[0.29, 0.27, 0.30]	[0.35, 0.27, 0.28]
B1-Exp.2 LF-LC	0.00	22.48	[0.17, 0.21, 0.12]	0.00	23.34	[0.17, 0.21, 0.12]	[0.17, 0.21, 0.12]
B1-Exp.2 LF-LC	**2.85**	23.91	[0.45, 0.44, 0.29]	**2.85**	24.74	[0.45, 0.44, 0.29]	[0.45, 0.51, 0.28]
YB-Exp. HS HD	0.00	20.97	[0.11, 0.10, 0.00]	0.00	21.71	[0.11, 0.10, 0.00]	[0.11, 0.10, 0.01]
YB-Exp. HS LD	0.00	20.97	[0.11, 0.05, 0.05]	0.00	21.71	[0.11, 0.04, 0.05]	[0.11, 0.04, 0.05]
YB-Exp. LS HD	0.00	20.97	[0.11, 0.05, 0.01]	0.00	21.71	[0.11, 0.05, 0.01]	[0.10, 0.05, 0.01]
YB-Exp. LS LD	0.00	20.97	[0.08, 0.06, 0.00]	0.00	21.71	[0.09, 0.06, 0.00]	[0.08, 0.05, 0.01]
Subtotal	27.11	277.57		22.44	284.91		
Three-source CPD task							
B1-Exp.3 HF-HC	**32.73**	50.48	[0.58, 0.65, 0.66, 0.29]	**31.23**	49.14	[0.62, 0.65, 0.63, 0.28]	[0.81, 0.57, 0.52, 0.29]
B1-Exp.3 HF-LC	**46.05**	57.16	[0.88, 0.74, 0.86, 0.49]	**31.45**	49.27	[0.84, 0.68, 0.94, 0.49]	[0.93, 0.54, 0.98, 0.49]
B1-Exp.3 LF-HC	**16.13**	42.18	[1.06, 0.87, 0.90, 0.25]	**16.61**	41.83	[1.10, 0.86, 0.86, 0.25]	[1.20, 0.86, 0.85, 0.24]
B1-Exp.3 LF-LC	**15.54**	41.88	[1.23, 0.94, 1.00, 0.43]	**13.00**	40.02	[1.24, 0.94, 0.97, 0.43]	[1.28, 0.87, 0.98, 0.44]
B1-Exp.4 HF-HC	**12.66**	37.10	[0.67, 0.49, 0.68, 0.35]	**11.59**	36.06	[0.74, 0.42, 0.68, 0.35]	[0.82, 0.46, 0.53, 0.36]
B1-Exp.4 HF-LC	**21.05**	41.27	[0.50, 0.61, 0.74, 0.54]	**19.43**	39.96	[0.51, 0.57, 0.79, 0.53]	[0.57, 0.53, 0.75, 0.54]
B1-Exp.4 LF-LC	5.49	33.51	[0.69, 0.64, 0.57, 0.19]	4.75	32.64	[0.72, 0.63, 0.55, 0.19]	[0.75, 0.59, 0.54, 0.19]
B1-Exp.4 LF-LC	2.91	32.22	[0.93, 0.68, 0.94, 0.31]	1.59	31.06	[0.92, 0.65, 0.96, 0.32]	[0.95, 0.59, 0.98, 0.32]
B2-Exp. HF-HC	**8.63**	36.87	[0.51, 0.56, 0.60, 0.31]	**8.68**	36.38	[0.54, 0.54, 0.60, 0.31]	[0.66, 0.49, 0.54, 0.31]
B2-Exp. HF-LC	1.54	33.32	[0.67, 0.64, 0.79, 0.47]	1.72	32.90	[0.68, 0.64, 0.79, 0.47]	[0.73, 0.60, 0.78, 0.47]
B2-Exp. LF-HC	0.52	32.97	[0.83, 0.62, 0.62, 0.20]	0.50	32.46	[0.83, 0.62, 0.62, 0.20]	[0.84, 0.63, 0.59, 0.20]
B2-Exp. LF-LC	4.29	34.13	[1.04, 0.82, 0.92, 0.36]	3.96	33.42	[1.05, 0.81, 0.91, 0.36]	[1.07, 0.74, 0.94, 0.36]
Subtotal	167.54	473.09		144.51	455.14		

Note: HF = high frequency, LF = low frequency, HC = high concreteness, LC = low concreteness, HS = high source-similarity, LS = low source-similarity, HD = high distractor-similarity, and LT = low distractor-similarity. B1 = Brainerd et al. (2012), B2 = Brainerd et al. (2013), and YB = Yu and Bellezza (2000). NML = normalized maximum likelihood. The critical G^2 value ($p = .05$) for the two-source data is 2.71 (sampling distribution follows a mixture of χ^2 distributions; see body of text) and for the three-source data is 7.81 (sampling distribution corresponds to $\chi^2_{df=3}$). G^2 results above the critical value are presented in bold. The parameter estimates (and confidence intervals) are provided in the Supplemental Material. The overdistribution values inside the squared brackets correspond (in order) to A, B, and new items in the case of the two-source CPD task, and to A, B, C, and new items in the case of the three-source CPD task.

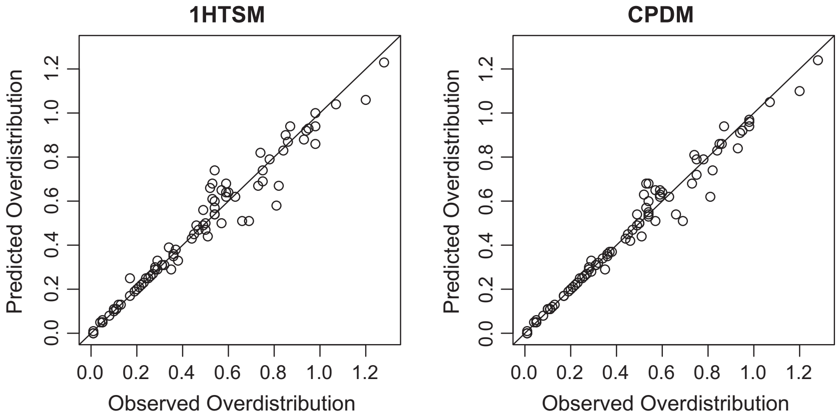

Fig. 3. Observed and predicted overdistribution values across the reanalyzed datasets.

then there is slight advantage for the 1HTSM (summed ΔNML = -2.78, $Z = 1.52, p = .14$).

A careful reevaluation of the two models shows that Brainerd et al.'s (2012) claims regarding the inability of the 1HTSM/2HTSM to account for the observed source overdistribution in comparison to the new CPDM were inaccurate. As it turns out, both models pretty much succeed and fail in similar circumstances. Furthermore, the claim that the particular order of processes in CPDM is meaningful was shown to be false. Altogether, these results suggest that the need for a new conceptualization of item and source memory processes (e.g., one that postulates a memory superposition state) is perhaps premature.

Still, the results are far from satisfactory given that both models often fail to provide good fits. Several reasons could be behind these large misfits; for example, the aggregation of individual data might produce distorted results (e.g., Heathcote, Brown, & Mewhort, 2000; Rouder & Lu, 2005; Rouder, Morey, & Pratte, in press). Another potential reason is that the models compared did not include any distractor-detection state, which is known to make an important contribution in accounting for the observed data (Bayen et al., 1996; Klauer & Kellen, 2010). Also, the use of sources with distinct memorability (i.e., sources being lists studied sequentially) might have led to the use of different retrieval strategies when attempting to remember whether an item belonged to a particular source, violating the assumption that memory processes (as captured by the parameters) are invariant across test probes (e.g., Marsh & Hicks, 1998).

In the next session we report two new experiments using an extended experimental design, providing more degrees of freedom for the full specification of models (e.g., a fully identifiable 2HTSM for two and three sources) and an evaluation of models that does not rely on aggregate data. One important aspect of this extension is that it capitalizes on some of the main assumptions of the models.

Source overdistribution: new experimental and modeling approaches

In this section we report two experiments that attempt to replicate source overdistribution in a setting that is closer to the source-memory experiments that are typically considered in the literature (e.g., DeCarlo, 2003; Hautus et al., 2008; Klauer & Kellen, 2010). An extended experimental design is proposed, allowing for the modeling of the 2HTSM as well as an extended version of the CPDM. Finally, the resulting data are modeled using hierarchical-Bayesian extensions of the models that avoid relying on sparse individual data or potentially distorted aggregate datasets (Rouder et al., in press).

Extending the response set in the CPD task

In the current CPD task, participants are required to answer "Yes" or "No" to each of the test probes. The use of a binary response set limits the ability to fully specify and estimate the parameters of the 2HTSM. For example, in a two-source experiment there are nine degrees of freedom available, which are insufficient to fully specify the ten free parameters of the 2HTSM. One way of increasing the number of degrees of freedom is by extending the response set, in this case by introducing a "Skip" response option which can be used to avoid a determinate (yes/no) response when uncertain. With a ternary response set, a two-source experiment provides 18 degrees of freedom.

The introduction of a skip response-option requires an extension of the models, which can be done in a rather straightforward manner: For the 2HTSM, the possibility of skipping is associated to cases in which the mental state does not provide enough information on whether the test item matches the test probe. For example, when only item memory is available (state M_3), individuals answer to probe "Source A?" by guessing "Yes", "Skip", and "No" with probabilities $\gamma_{y|A?}$, $(1-\gamma_{y|A?})\gamma_{s|A?}$, and $(1-\gamma_{y|A?})(1-\gamma_{s|A?})$, respectively. Different guessing parameters are specified for the "Source B?" probe ($\gamma_{y|B?}$ and $\gamma_{s|B?}$). Note that no parameters have to be specified for the "Old?" probe given that item memory is sufficient to respond with certainty in state M_3. When no memory is available at all (state M_5), individuals answer to probe "Source A?" by guessing "Yes", "Skip", and "No" with probabilities $\beta_{y|A?}$, $(1-\beta_{y|A?})\beta_{s|A?}$, and $(1-\beta_{y|A?})(1-\beta_{s|A?})$, respectively. Different guessing parameters are specified for probes "Source B?" ($\beta_{y|B?}$ and $\beta_{s|B?}$) and "Old?" ($\beta_{y|O?}$ and $\beta_{s|O?}$). Altogether, the 2HTSM has fifteen free parameters, all of them fully identifiable. This type of extension has been previously proposed and implemented by Oravecz, Faust, and Batchelder (in press) and Singmann, Kellen, and Klauer (2013). To better illustrate this extension, Fig. 4 specifies the 2HTSM's tree for A items with the ternary response set.

According to the theoretical principles underlying the CPDM, "Skip" responses should only occur when no memory is available (state W_6). This simply follows from the fact that all other memory states deterministically lead to either "Yes" or "No" responses, depending on the state. This means that the model is extended by having additional β parameters in the exact same way as in the 2HTSM. The resulting CPDM has twelve free parameters, all of them fully identifiable. See Fig. 4 for a depiction of the CPDM's tree for A items with the ternary response set.

It is important to note that at the core of the CPDM there are extremely strong theoretical assumptions regarding the mapping of memory states onto observed responses (i.e., these restrictions do not result from a pragmatic choice due to modeling constraints). In fact this response determinism plays a fundamental role when arguing that source overdistribution is the outcome of an item-memory (or superposition) state that is *invariably* mapped onto "yes" responses, *independently* of the test probe.

The remaining degrees of freedom allow for the CPDM to be further extended, by introducing an additional state (W_7) where a distractor item is detected as being new (with probability D_N), equivalent to M_4 in the 2HTSM. This extended model, henceforth referred as CPDM$_N$, is included among the candidate models in our analysis. Including this extended model prevents the possibility of preferring the 2HTSM to the CPDM simply because the latter does not include a distractor-rejection mechanism, whereas the former does.

Hierarchical-Bayesian parameter estimation and model comparison

One of the common issues in cognitive modeling concerns the choice between individual and group data. It is well known that data aggregation glosses over participant heterogeneity and can lead to severe distortions when relying on non-linear models (e.g., Estes & Maddox, 2005; Heathcote et al., 2000; Rouder et al., in press). This suggests that a more accurate characterization of the underlying cognitive processes can be achieved by relying on individual data. On the other hand, the reliance on individual-level data can be equally problematic when the number of observations per individual is small, as such data are likely to produce noisy and unreliable parameter estimates (e.g., Chechile, 2009; Cohen, Sanborn, & Shiffrin, 2008).

A principled compromise between both approaches is given by considering hierarchical extensions of the models (Cohen et al., 2008; Rouder et al., in press): It is assumed that individual parameters originate from group-level distributions, such as a (multidimensional) Gaussian distribution. The standard deviations of the group-level distribution reflect the existing differences between participants, being small when participants are rather homogeneous, and large when there are substantial individual differences. The hierarchical approach allows for the individual differences and similarities to be captured in a single, principled framework, overcoming the limitations of both individual and group-level data (for comprehensive introductions to hierarchical modeling,

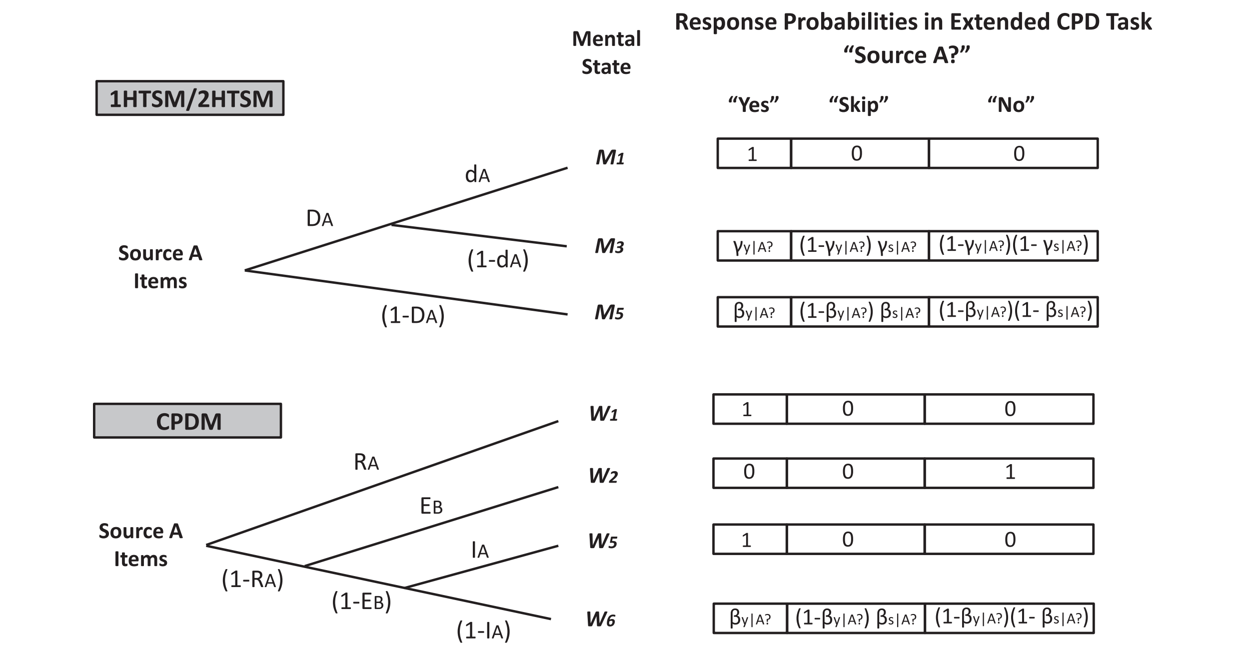

Fig. 4. 1HTSM/2HTSM and CPDM trees for Source A items and a ternary response set (test probe "Source A?" only). Note that each set of response probabilities associated to a mental state sums to 1.

see Lee & Wagenmakers, 2013; Rouder & Lu, 2005; Rouder et al., in press).

Because of its advantages, hierarchical extensions of cognitive models have been regularly developed and proposed across the recent literature (e.g., Lee & Wagenmakers, 2013). Such developments have especially been prominent in the MPT model class (Klauer, 2006, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2013; Rouder, Lu, Morey, Sun, & Speckman, 2008; Smith & Batchelder, 2010), a class to which both the 2HTSM and the CPDM are members of.

In order to see how the hierarchical extension is introduced, let $\Theta : \theta_{i,j}, i = 1, \ldots, I, j = 1, \ldots, J$ represent the I sets of J individual parameters in a given (MPT) model. For the ith participant, the jth parameter is given by

$$\theta_{i,j} = \Phi(\mu_j + \xi_j \delta_{i,j}) \tag{8}$$

where Φ is the standard normal distribution function, μ_j is the group mean of parameter $\theta_{i,j}$, and $\delta_{i,j}$ is the individual displacement from the group mean. Parameters $\delta_{i,j}$ are draws from a zero-centered multivariate Gaussian distribution $\mathcal{N}(0, \Sigma)$. The covariance matrix Σ captures any potential correlations between parameters across individuals (Klauer, 2010). Parameter ξ_j is a redundant multiplicative scale-parameter that only serves to accelerate the rate of convergence in the Markov chain Monte Carlo (MCMC) sampling method used for parameter estimation under a Bayesian framework (see Gelman, Carlin, Stern, & Rubin, 2004, chap. 15; Klauer, 2010).[7]

Hierarchical extensions of models can be implemented in a Bayesian framework (e.g., Gelman & Hill, 2007), whereas their implementation under a classic-frequentist framework is difficult (see Klauer, 2010). In a Bayesian framework, the available information regarding a model and its parameters is represented by probability distributions. In particular, *posterior* distributions are obtained by using the observed data to update established *prior* distributions. These priors can incorporate knowledge obtained from previous findings, or alternatively be rather uninformative, as in the present case. The following prior distributions are associated to each parameter in the present application (following Klauer (2010), Matzke et al. (2013), & Rouder et al. (2008)):

$$\mu_j \sim \mathcal{N}(0, 1),$$
$$\xi_j \sim \mathcal{N}(0, 1),$$
$$\Sigma \sim \mathcal{W}^{-1}(Id(J), J + 1),$$

where $\mathcal{N}$ is the Gaussian distribution and $\mathcal{W}^{-1}(Id(J), J + 1)$ is the so-called inverse Wishart distribution with $J + 1$ degrees of freedom, with $Id(J)$ being the identity matrix with J rows and columns (see Gelman & Hill, 2007, chap. 13).

In order to assess the fit of each model, we rely on so-called posterior model checks using Bayesian p-values (see Gelman et al., 2004, chap. 6). Bayesian p-values will be computed for the T_1 and T_2 test statistics proposed by Klauer (2010): Statistic T_1 quantifies a model's ability to account for the total observed category frequencies, aggregated across individuals, and T_2 the ability of a model to account for the variances and covariances in the observed category frequencies (for details on the T_1 and T_2 test statistics, see Klauer (2010)). The model's posterior distributions can also be used to produce predictions concerning overdistribution.

Finally, due to the current inability to compute NML for hierarchical implementations, the comparison between the candidate models is done using the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002), a model-selection index that generalizes the Akaike and Bayesian information criteria (AIC and BIC; see Burnham & Anderson, 2002) to hierarchical models estimated using MCMC methods. The DIC penalizes models by taking into account their respective number of *effective* parameters (denoted by p_D), which are directly estimated from the MCMC chains. The number of effective parameters in a model corresponds to the number of parameters in a model that are not constrained by the imposed hierarchical structure (for details, see Spiegelhalter et al., 2002). Like for AIC and BIC, the model with the lowest DIC value is the one striking the best trade-off between goodness of fit and parsimony.

Experiment 1

The goal of the first experiment is to obtain a source-memory dataset with ternary responses that can be used for comparing the hierarchical extensions of the 2HTSM, CPDM, and $CPDM_N$. In this experiment, two sources were used, which differed in their visual characteristics (color and spatial position).

Method

Participants

32 undergraduate psychology students (mean age = 21.5, SD = 3.18, ranging from 18 to 30 years) from the University of Freiburg served as participants for the experiment. They received course credits for their participation. Subjects were native speakers of German without severe or uncorrected visual impairments. Each participant was tested individually, and one session of the experiment lasted approximately 30 min.

Design and procedure

The computer-based experiment consisted of a single study phase followed by a single test phase. Prior to the study phase, individuals were informed about the two sources and their defining characteristics. In the study phase 180 words (90 A and 90 B items) were presented in a randomized order. Each word was presented for 2000 ms, with 500 ms inter-stimulus interval. Source A items were presented in red on the left side of the screen

[7] The items used in the experiments reported below were selected anew from a larger pool for each participant. This item sampling scheme considerably reduced the number of "replications" of each item across all levels of the experiment design (item type × test probe), compromising any realistic attempt to test and/or model item heterogeneity along with participant heterogeneity (see Matzke et al., 2013).

and B items in blue on the right side of the screen, against a light gray background. The assignment of words to each source was randomized anew for each participant.

Prior to the test phase, participants were informed about the different test probes and the relative proportion of items in the test list ($\frac{1}{3}$ A items, $\frac{1}{3}$ B items, and $\frac{1}{3}$ New items). Furthermore, participants were reminded that no item could simultaneously belong to both source A and B. To encourage participants to provide accurate responses as well as consider the use of the "Skip" response, they were informed that they would receive 1 point per correct response and lose 3 points per incorrect responses, and that they could use the "Skip" response to avoid losing points when uncertain about the true status of the item. Also, it was stated that the participant with the best score would receive a €30 prize. In the test phase, a total of 270 items were presented (90 A items, 90 B items and 90 new items). All words were presented in black at the center of the screen, against a light gray background. One third of each item type was tested with one of the three test probes ("Source A?", "Source B?", and "Old?"). The test probe was presented above the test item, with "Source A?"and "Source B?" presented in red and blue, respectively. Probe "Old?" was presented in black. Three response buttons were presented below the test word, and participants could select a response by clicking on it using a mouse. No feedback was given throughout the test, with the individual's final score being presented at the end of the test phase. After completing the test phase, participants were thanked and debriefed.

Materials

The experiment was implemented in PsychoPy (Peirce, 2007). Words were sampled from a selection of 639 words from Lahl, Göritz, Pietrowsky, and Rosenberg (2009), ranging from 4 to 8 letters in length. According to the ratings obtained by Lahl et al., the words were all of medium valence (ranging from 3.50 to 6.50 on an 11-point scale) and low in arousal (ranging from 0.50 to 4.50 on an 11-point scale). Furthermore, all words were of approximately equal word frequency, as indicated by the log frequency ratings obtained for each word via WordGen (ranging from 0.30 to 2.90; Duyck, Desmet, Verbeke, & Brysbaert, 2004).

Results and discussion

The average proportion of overall correct responses was .67 (excluding "Skip" responses) and the average proportion of "Skip" responses was .26. The average source overdistribution for A, B, and new items was .30, .29, and .17, respectively, in line with values reported by Brainerd et al. (2012). Overdistribution was found to be reliably above zero (smallest $t(31) = 4.68$, largest $p < 0.01$).

Prior to the modeling analysis, it is important to test the data for participant homogeneity (Smith & Batchelder, 2008). A χ^2 test of participant homogeneity was used separately for each multinomial distribution in the data (each corresponding to a item type × test probe combination). In each multinomial distribution the participant-homogeneity hypothesis was rejected (smallest $\chi^2(62) = 202.36$, largest $p < .001$). A bootstrap analysis that does not rely on asymptotic approximations confirms that the results are statistically significant. Following Smith and Batchelder (2008), this rejection strongly discourages the use of aggregate data and recommends the use of a hierarchical model that can account for participant heterogeneity.

Modeling results are presented in Table 2[8]: Contrary to what is claimed by Brainerd et al. (2012), the 2HTSM provided a suitable characterization of the data, as quantified by the p-values of the T_1 and T_2 statistics. This result was further corroborated by the predicted average source-overdistribution values that were very close to the observed ones. In contrast, CPDM failed to account for the data, as indicated by the T_1 statistic for aggregate frequencies and the T_2 statistic for the variance–covariance structure (all $p < .05$), and also failed to provide accurate mean posterior-predictive estimates of source overdistribution. Mean posterior-predictive overdistribution values are obtained by sampling parameters from a model's posterior distributions and then generating data using these parameters. This two-step sampling process is repeated multiple times and then averaged. On the other hand, CPDM$_N$ improved upon the results, supporting the inclusion of a distractor-rejection state in the model. This was observed in the non-significant p-values associated to the T_1 and T_2 statistics, and the mean posterior-predictive overdistribution values that were closer to the observed ones.

Regarding the relative performance of the candidate models, the 2HTSM had a lower DIC than the CPDM (ΔDIC = 166.91) and the CPDM$_N$ (ΔDIC = 52.90). According to the DIC metric, which can be interpreted in the same way as AIC and BIC (Burnham & Anderson, 2002), differences larger than 10 represent extremely strong evidence in favor of the winning model, in this case the 2HTSM.

Concerning the parameter estimates, we will focus on the 2HTSM and CPDM$_N$. The group-mean parameter estimates of the models are reported in Table 2, along with their respective 95% credible intervals, which can be interpreted much like 95% confidence intervals. The parameters of the 2HTSM show that source memory (d_A and d_B) was low, especially when compared to the overall item memory and distractor rejection (D_A, D_B, and D_N). Parameter correlations show a strong relationship between item detection for the two sources ($\rho(D_A, D_B) = 0.77$ [0.37, 0.93]) that was expected given the equal memorability of items from the two sources. Still, no reliable correlations were found for source memory ($\rho(d_A, d_B) = 0.62$ [−0.37, 0.93]) nor between item memory and distractor rejection ($\rho(D_A, D_N) = 0.19$ [−0.36, 0.65] and $\rho(D_B, D_N) = 0.14$

[8] The hierarchical models were implemented in C++ along with the NAG library, using the scripts developed by Klauer (2010). Scripts can be obtained from the authors upon request. Four independent streams of MCMC samples were collected for each model using a Gibbs sampler. Rough initial estimates of the parameters were obtained by means of the Monte Carlo EM algorithm (Wei & Tanner, 1990). Chain convergence was assessed via the $\hat{R}$ statistic, which compares within-chain variance to between-chain variance (Gelman et al., 2004, chap. 11). Sampling with the Gibbs sampler continued until all MCMC streams converged (all $\hat{R} \leqslant 1.05$), and then went on for 1000 consecutive samples per stream, for a total of 4,000 draws from the posterior parameter distributions retained for analysis.

Table 2
Model-Fitting Results for Experiment 1.

				2HTSM			
Model fit	T_1	T_2	p_D	DIC	ΔDIC		
	.44	.52	233.36	15165.60	0		
Parameter estimates	D_A	D_B	D_N	d_A	d_B		
	.52 [.44, .63]	.51 [.40, .63]	.25 [.16, .38]	.11 [.01, .25]	.16 [.05, .34]		
	$\beta_{y\|A?}$	$\beta_{s\|A?}$	$\beta_{y\|B?}$	$\beta_{s\|B?}$	$\beta_{y\|O?}$	$\beta_{s\|O?}$	
	.18 [.12, .25]	.65 [.41, .85]	.26 [.16, .38]	.66 [.43, .91]	.25 [.18, .34]	.33 [.21, .44]	
	$\gamma_{y\|A?}$	$\gamma_{s\|A?}$	$\gamma_{y\|B?}$	$\gamma_{s\|B?}$			
	.73 [.53, .91]	.11 [.00, .55]	.75 [.52, .90]	.09 [.01, .76]			
Predicted overdistribution	Source A	Source B	New				
	0.28	0.27	0.17				
				CPDM			
Model fit	T_1	T_2	p_D	DIC	ΔDIC		
	< .01	.03	213.09	15332.51	166.91		
Parameter estimates	R_A	R_B	E_A	E_B	I_A	I_B	
	.10 [.04, .18]	.13 [.06, .21]	.01 [.00, .06]	.04 [.00, .10]	.36 [.26, .46]	.33 [.23, .43]	
	$\beta_{y\|A?}$	$\beta_{s\|A?}$	$\beta_{y\|B?}$	$\beta_{s\|B?}$	$\beta_{y\|O?}$	$\beta_{s\|O?}$	
	.12 [.07, .18]	.54 [.40, .72]	.17 [.11, .24]	.54 [.39, .74]	.21 [.15, .29]	.30 [.19, .46]	
Predicted overdistribution	Source A	Source B	New				
	0.39	0.36	0.12				
				CPDM_N			
Model fit	T_1	T_2	p_D	DIC	ΔDIC		
	.35	.09	222.57	15218.50	52.90		
Parameter estimates	R_A	R_B	E_A	E_B	I_A	I_B	D_N
	.17 [.07, .24]	.19 [.13, .26]	.12 [.06, .19]	.10 [.04, .19]	.29 [.19, .40]	.26 [.10, .36]	.26 [.16, .37]
	$\beta_{y\|A?}$	$\beta_{s\|A?}$	$\beta_{y\|B?}$	$\beta_{s\|B?}$	$\beta_{y\|O?}$	$\beta_{s\|O?}$	
	.18 [.12, .29]	.67 [.48, .83]	.28 [.18, .40]	.69 [.51, .89]	.27 [.20, .36]	.31 [.18, .42]	
Predicted overdistribution	Source A	Source B	New				
	0.32	0.30	0.15				

Note: DIC is computed up to an additive constant. ΔDIC corresponds to the difference between the model's DIC and the DIC of the best-performing model. The parameter estimates correspond to the group-mean parameters estimated with the hierarchical models. The values between squared brackets are the 95% credible intervals. The values under T_1 and T_2 correspond to the *p*-values of these statistics. The predicted overdistribution values are the averages obtained from each model's sampled predictive-posterior distributions. For reference, the average source overdistribution values observed for A, B, and new items were .30, .29, and .17, respectively.

[−0.44, 0.62]) given that the 95% credible intervals include the value 0 in these cases. Reliable correlations between the guessing parameters were also found for most guessing parameters, indicating that individuals behaved in a similar manner for the different probes (e.g., $\rho(\beta_{y|A?}, \beta_{y|B?}) = 0.71$ [0.23, 0.92], $\rho(\beta_{y|A?}, \beta_{y|O?}) = 0.72$ [0.27, 0.91], and $\rho(\gamma_{y|A?}, \gamma_{y|B?}) = 0.13$ [−0.60, 0.78]), and that individuals manifest a general tendency to skip ($\rho(\beta_{s|A?}, \beta_{s|B?}) = 0.90$ [0.68, 0.97], $\rho(\beta_{s|A?}, \beta_{s|O?}) = 0.62$ [0.13, 0.85], $\rho(\gamma_{s|A?}, \gamma_{s|B?}) = 0.73$ [0.06, 0.94]). Full estimates of each model's variance–covariance matrix can be found in the Supplemental Material.

Let us now turn to the CPDM_N: Concerning true recollection, a positive relationship was found, although not a reliable one ($\rho(R_A, R_B) = 0.54$ [−0.07, 0.88]). No relationship was found for false recollection ($\rho(E_A, E_B) = 0.07$ [−0.86, 0.94]). On the contrary, a reliable positive correlation was found between the item memory parameters ($\rho(I_A, I_B) = 0.74$ [0.21, 0.94]) as well as between the guessing parameters ($\rho(\beta_{y|A?}, \beta_{y|B?}) = 0.72$ [0.27, 0.91]). Also, a general tendency for skipping was reliably found ($\rho(\beta_{s|A?}, \beta_{s|B?}) = 0.90$ [0.66, 0.97], $\rho(\beta_{s|A?}, \beta_{s|O?}) = 0.62$ [0.18, 0.86]).

The results from Experiment 1 show that overdistribution can be described as a byproduct of different guessing processes, without the need of assuming additional memory states such as false recollection or a memory state that invariably leads to "Yes" responses. The model proposed by Brainerd et al. (2012) failed to provide a good account of the data, as shown by the different model-performance statistics used. An extension of Brainerd et al.'s model performed considerably better but still fared worse than the 2HTSM.

Besides accounting for overdistribution, the 2HTSM also provided a good characterization of the different underlying processes as well as the relationships between them, as shown by the estimated parameter correlations. Still, no reliable positive correlation was found between the source-memory parameters in the 2HTSM (although a positive relationship was found). One factor contributing to this result may be the fact that source-memory was generally low, which can compromise the detection of correlations in source memory. Another reason for the non-reliable correlation may be the small number of participants in this experiment. The low source-memory performance could also explain the lack of reliable correlations between recollection parameters in CPD_N. It should be noted though that the low source-memory performance that was observed is not problematic per se given that overdistribution is expected to be large when there is little recollection of the sources (see Brainerd et al., 2012, Figure 2). In fact, if the phenomenon of overdistribution is capable of distinguishing between the competing models, then the present case should be most informative. Nevertheless, a rigorous comparison between the different models should include cases in which source-memory performance is somewhat higher. One of the benefits would be a greater similarity to previously published studies (e.g., DeCarlo, 2003; Hautus et al., 2008; Klauer & Kellen, 2010).

Experiment 2

The purpose of Experiment 2 is to attempt to replicate and generalize the findings of Experiment 1. First, following Brainerd et al.'s (2012) experiments 3 and 4, three sources are used instead of only two. The extension of the models to a three-source experiment is rather straightforward and therefore will not be described here (but model equations can be found in the Supplemental Material). Second, the study phase was modified in order to improve source-discrimination and place the overall performance closer to what is usually observed in the source-memory literature (e.g., Klauer & Kellen, 2010). With the exception of the details described below, this study was implemented in the exact same way as Experiment 1.

Method

Participants

42 individuals (35 university students; mean age = 24, SD = 3.33, ranging from 19 to 35 years) served as participants for the experiment. In exchange for their participation, participants received between €5 and €8, depending on their performance. Each participant was tested individually, and the experiment lasted approximately 40 min.

Design and procedure

The study phases consisted of two study blocks in which participants learned 240 words (80 A, 80 B, and 80 C items). In both study blocks all 240 items were presented, with a different randomized order per block. Additionally, the same sets of six primacy and six recency buffers (two for each source) were presented before and after each block, with a different randomized order per block. Each word was presented for 1200 ms, with 500 ms inter-stimulus interval. Before the first study block, participants were made aware that there would be two study blocks with identical items.

Source A items were presented in red on the left side of the screen, B items in blue on the top, and C items in green on the right side, against a light gray background. The vertical position of A and C items was slightly lowered in order to have the positions of A, B, C items equidistant from each other and from the center of the screen (i.e., in a equilateral triangle with the screen center being the center of the triangle). The assignment of words to the sources was randomized anew for each participant. A participant's total score (scored as in Experiment 1) from the test phase was converted into Euro cents so that a perfect score would yield a payoff of €3 that was added to a fixed compensation of €5. No feedback was given throughout the test phase.

Results and discussion

The average proportion of overall correct responses was .71 (excluding "Skip" responses) and the average proportion of "Skip" responses was .38. Mean overdistribution values for A, B, C, and new items were 0.34, 0.40, 0.37, and 0.19, respectively, with overdistribution being reliably above zero for all item types (smallest $t(41) = 3.19$, largest $p < .01$). As in Experiment 1, we tested the assumption of participant homogeneity: In each multinomial distribution the participant-homogeneity hypothesis was rejected (smallest $\chi^2(82) = 386.10$, largest $p < .001$), a result that was confirmed by a bootstrap analysis. Once again, the use of a hierarchical-modeling approach is found to be justified.

In terms of model performance, the results were similar to the ones obtained in Experiment 1: As shown in Table 3, the 2HTSM provided a good account for the data, followed by the $CPDM_N$. Again the CPDM failed to account for the data ($p < .01$ for both T_1 and T_2). The same pattern was found in the DIC results, with the 2HTSM performing better than the CPDM (ΔDIC = 210.66) and the $CPDM_N$ (ΔDIC = 43.45). One way in which the superiority of the 2HTSM manifested itself was in terms of the predicted overdistribution values, which were much closer to the actual values than the predictions made by the CPDM, although the $CPDM_N$ also predicted overdistribution values that were relatively close to the observed ones.

Regarding the 2HTSM's parameter estimates (see Table 3), it is clear that, in comparison with Experiment 1, source-memory parameter estimates increased while item-memory parameters slightly decreased, leading to overall parameter estimates that are close to what is usually found in the literature (e.g., Klauer & Kellen, 2010). These differences in opposite directions in comparison to Experiment 1 can be seen as strange at first glance given that the words were presented twice in Experiment 2. However, this difference is not unexpected given that the words in Experiment 2 were also presented *faster*, therefore not guaranteeing an increase of item-memory, and the block-repetition scheme used reduced the average temporal lag between item study and test, a lag which is known to disproportionately impair source memory (e.g., Jennings & Jacoby, 1997). Reliable correlations were now found between the item and the source-memory parameters ($\rho(D_A, D_B) = 0.83$ [0.61, 0.94], $\rho(D_B, D_C) = 0.82$ [0.55, 0.93], $\rho(d_A, d_B) = 0.78$ [0.37, 0.94], and $\rho(d_B, d_C) = 0.77$ [0.37, 0.94]), but not between distractor rejection and item

Table 3
Model-Fitting Results for Experiment 2.

				2HTSM				
Model fit	T_1	T_2	p_D	DIC	ΔDIC			
	.06	.11	392.08	21216.01	0			
Parameter estimates	D_A	D_B	D_C	D_N	d_A	d_B	d_C	
	.43 [.30, .53]	.47 [.34, .58]	.39 [.23, .50]	.15 [.06, .23]	.45 [.29, .62]	.45 [.25, .61]	.51 [.36, .65]	
	$\beta_{y\|A?}$	$\beta_{s\|A?}$	$\beta_{y\|B?}$	$\beta_{s\|B?}$	$\beta_{y\|C?}$	$\beta_{s\|C?}$	$\beta_{y\|O?}$	$\beta_{s\|O?}$
	.10 [.03, .18]	.68 [.52, .81]	.11 [.05, .19]	.65 [.43, .84]	.15 [.08, .23]	.78 [.60, .94]	.19 [.11, .28]	.43 [.25, .60]
	$\gamma_{y\|A?}$	$\gamma_{s\|A?}$	$\gamma_{y\|B?}$	$\gamma_{s\|B?}$	$\gamma_{y\|C?}$	$\gamma_{s\|C?}$		
	.69 [.51, .84]	.69 [.21, .99]	.78 [.62, .98]	.79 [.43, .97]	.71 [.38, .84]	.69 [.29, .94]		
Predicted overdistribution	Source A	Source B	Source C	New				
	0.35	0.37	0.33	0.21				
				CPDM				
Model fit	T_1	T_2	p_D	DIC	ΔDIC			
	< .01	< .01	370.12	21426.67	210.66			
Parameter estimates	R_A	R_B	R_C	E_A	E_B	E_C		
	.17 [.09, .26]	.19 [.10, .29]	.15 [.08, .26]	.00 [.00, .01]	.00 [.00, .01]	.00 [.00, .01]		
	I_A	I_B	I_C					
	.16 [.08, .26]	.18 [.09, .29]	.16 [.09, .26]					
	$\beta_{y\|A?}$	$\beta_{s\|A?}$	$\beta_{y\|B?}$	$\beta_{s\|B?}$	$\beta_{y\|C?}$	$\beta_{s\|C?}$	$\beta_{y\|O?}$	$\beta_{s\|O?}$
	.06 [.03, .10]	.63 [.45, .77]	.07 [.03, .11]	.60 [.40, .74]	.09 [.06, .14]	.69 [.50, .84]	.17 [.11, .25]	.43 [.22, .61]
Predicted overdistribution	Source A	Source B	Source C	New				
	0.40	0.41	0.39	0.14				
				CPDM_N				
Model fit	T_1	T_2	p_D	DIC	ΔDIC			
	.05	.01	363.63	21259.46	43.45			
Parameter estimates	R_A	R_B	R_C	E_A	E_B	E_C		
	.19 [.11, .30]	.22 [.15, .30]	.18 [.11, .28]	.01 [.00, .02]	.01 [.00, .04]	.01 [.00, .04]		
	I_A	I_B	I_C	D_N				
	.13 [.03, .20]	.15 [.10, .23]	.12 [.08, .18]	.18 [.10, .29]				
	$\beta_{y\|A?}$	$\beta_{s\|A?}$	$\beta_{y\|B?}$	$\beta_{s\|B?}$	$\beta_{y\|C?}$	$\beta_{s\|C?}$	$\beta_{y\|O?}$	$\beta_{s\|O?}$
	.11 [.05, .19]	.66 [.51, .91]	.11 [.07, .20]	.64 [.44, .91]	.14 [.06, .25]	.80 [.58, .94]	.23 [.13, .32]	.43 [.26, .61]
Predicted overdistribution	Source A	Source B	Source C	New				
	0.37	0.39	0.37	0.17				

Note: DIC is computed up to an additive constant. ΔDIC corresponds to the difference between the model's DIC and the DIC of the best-performing model. The parameter estimates correspond to the group-mean parameters estimated with the hierarchical models. The values between squared brackets are the 95% credible intervals. The values under T_1 and T_2 correspond to the *p*-values of these statistics. The predicted overdistribution values are the averages obtained from each model's sampled predictive-posterior distributions. For reference, the average source overdistribution values observed for A, B, C, and new items were 0.34, 0.40, 0.37, and 0.19, respectively.

memory (e.g., $\rho(D_N, D_A) = 0.22$ [−0.24, 0.62]). In terms of guessing parameters the results were similar to Experiment 1, with strong reliable correlations being found between the guessing parameters (e.g., $\rho(\beta_{y|A?}, \beta_{y|B?}) = 0.86$ [0.59, 0.96]) as well as a tendency to skip responses (e.g., $\rho(\beta_{s|A?}, \beta_{s|B?}) = 0.87$ [0.68, 0.96] and $\rho(\beta_{s|A?}, \beta_{s|O?}) = 0.65$ [0.30, 0.84]).

Concerning the CPDM_N's parameter values, a strong reliable correlation was found between the true-recollection parameters (e.g., $\rho(R_A, R_B) = 0.84$ [0.58, 0.95]) and the item-memory parameters (e.g., $\rho(I_A, I_B) = 0.79$ [0.38, 0.94]), but not for the false-recollection parameters (e.g., $\rho(E_A, E_B) = 0.11$ [−0.79, 0.91]). One reason for the latter result were the extremely low values taken on by the *E* parameters. Given the low *E* parameter estimates along with the increase of the *R* estimates (in comparison with Experiment 1), there is the possibility of a negative relationship between *R* and *E*. None of the estimates suggest so (e.g., $\rho(R_A, E_A) = -0.16$ [−0.88, 0.87] and $\rho(R_A, E_B) = -0.01$ [−0.81, 0.81]), but the level of uncertainty in these estimates is large. As in Experiment 1, strong reliable correlations were found between the guessing parameters (e.g., $\rho(\beta_{y|A?}, \beta_{y|B?}) = 0.87$ [0.64, 0.97]) as well as a tendency to skip responses (e.g., $\rho(\beta_{s|A?}, \beta_{s|B?}) = 0.88$ [0.71, 0.96] and $\rho(\beta_{s|A?}, \beta_{s|O?}) = 0.65$ [0.34, 0.84]).

Experiment 2 replicated the main finding of Experiment 1: Not only was source overdistribution accounted for by the 2HTSM, this model also provided a better account than Brainerd et al.'s (2012) candidate model.

General discussion

The CPD task extended Jacoby's (1991) original Process Dissociation task, enabling the estimation of several contributing processes in the recognition of semantically-related lists (Brainerd et al., 1999) as well as the observation of emerging phenomena like memory overdistribution (Brainerd & Reyna, 2008). The recent use of the CPD task in source-memory studies revealed the presence of source overdistribution, which according to Brainerd et al. (2012) has wide implications in the understanding and modeling of source-memory judgments. Instead of the traditional distinction between (a) (veridical) item memory, (b) (veridical) source memory dependent on item memory, and (c) guessing (Hautus et al., 2008; Klauer & Kellen, 2010; Onyper et al., 2010), Brainerd et al. (2012) proposed a new model characterizing source-memory judgments through the distinction between (a) true source-recollection, (b) false source-recollection, (c) item memory (superposition), and (d) guessing. Two important aspects were associated to this new characterization: First, the model assumed a quantum-like item-memory/superposition state (Brainerd et al., 2013). Second, the apparent independence of the different source-recollection and item memory processes was claimed to be meaningful, with source memory being retrieved in the absence of item memory. These two aspects represent theoretical departures to the way source memory is usually modeled.

The present work argued that a traditional source-memory model like the 1HTSM/2HTSM is able to account for source overdistribution via its different guessing processes. Subsequently, we evaluated how well these two different accounts, the 1HTSM/2HTSM and the CPDM, were able to account for source overdistribution. A reanalysis of both previously-published data as well as new experimental results using an extended task indicated that the observed source overdistribution could be attributed to a set of rather mundane and well studied guessing processes (e.g., Riefer et al., 1994; Schütz & Bröder, 2011) and the well-known phenomenon of response conservatism (e.g., Thomas & Legge, 1970). This guessing-based account is in line with traditional modeling accounts of source memory and runs counter to the claims of Brainerd et al. (2012, 2013) that a new conceptualization of memory processes is necessary. Instead of leading to a dismissal of traditional models of item and source memory, source overdistribution proved to be well accounted for by the 2HTSM, restating the value of this well-known model in decomposing observed responses into basic cognitive processes.

The two experiments reported here are in some ways distinct from the ones reported by Brainerd et al. (2012, 2013): (1) The sources correspond to different colors and spatial positions instead of lists studied sequentially; (2) a set of homogeneous items was used instead of different types of items (e.g., low versus high frequency); (3) a "Skip" response option was made available. Given such differences, one could argue that despite the fact that source overdistribution is observed in the present experimental results, they are not comparable to the results by Brainerd et al. (2012, 2013). Several reasons lead us to think otherwise: First, source overdistribution is presented by Brainerd et al. (2012) as a phenomenon that is informative regarding the nature of the memory processes underlying source-memory performance. If that is the case, then one would expect such phenomenon to play a role in source memory performance in general, such as when homogeneous words are tested and the sources only differ in terms of simple perceptual features (a common choice in source-memory studies; e.g., Batchelder & Riefer, 1990). Second, the introduction of a "Skip" option should make it easier to distinguish between the 1HTSM/2HTSM and CPDM given the *theory-driven*, *deterministic* state-response mapping of the CPDM, which should exclude the use of the "Skip" option in many instances. In fact, the distinction between the 1HTSM/2HTSM and the CPDM could be further explored by considering a manipulation that selectively influenced the use of the "Skip" option (e.g., a payoff manipulation). Another possibility would be to provide participants with feedback during test in order to encourage them to approximate their guessing tendencies to the test-item base rates. According to the 1HTSM/2HTSM's pure guessing account, this approximation should reduce or even eliminate overdistribution.

Of course, one could in principle relax the CPDM mappings when a "Skip" option is available but it is then not clear why such a relaxation was not already considered for the deterministically mapped "No" responses, and more importantly, how it affects the theoretical status of the memory states themselves. For example, if the item-memory/superposition state of the CPDM (state W_5) is not deterministically mapped onto "Yes" responses but can also be mapped onto "Skip" or "No" responses, then it becomes equivalent to the item-memory state of the 1HTSM/2HTSM (e.g., state M_3) in which responses are produced via guessing. As previously shown, such equivalences would make the two models almost indistinguishable as they can be seen as special cases of an encompassing model.

The existence of an encompassing model also offers interesting opportunities for future work, especially with the development of richer experimental designs that allow its identifiability and testability (e.g., tests of selective influence). Such designs might include both source-memory and CPD tasks (e.g., Yu & Bellezza, 2000), larger sets of test probes (e.g., "both Source A and B?"; see Brainerd et al., in press), multiple study conditions (e.g., Brainerd et al., in press; Gallo, Weiss, & Schacter, 2004), among other possibilities. Additionally, efforts can also be placed on better understanding item effects and the impact of item features such as word frequency within a hierarchical-Bayesian framework (Matzke et al., 2013).[9] This approach has the advantage of sidestepping item-aggregation effects (e.g., classifying items as "high frequency" and "low frequency") and providing a more fine-grained

[9] One potential source of distortion that could be removed via the introduction of item effects or the use of item features as covariates is the role of semantic/associative relatedness. Although the sources in both experiments were not associated with any particular group or category, it is possible that the association between single items (e.g., FLOWER-PLANT) could have led to some unaccounted effects (e.g., distractor PLANT being associated to the source in which FLOWER was studied).

understanding (see Rouder et al., in press). The same applies for individual-related variables such as age, which are known to be associated with important differences at the level of the memory processes (e.g., Ceci et al., 2010).

Finally, consider the implications of the present results for the broader literature. The present work focused on source memory, contrary to the original work on episodic overdistribution which focused on lists of semantically-related words (Brainerd & Reyna, 2008). As noted by Stahl and Klauer (2008) in a series of validation studies, the CPDM seems to provide a more adequate account of the latter kind of data than the 1HTSM. This advantage of the CPDM for the case of semantic lists stems from the fact that it is a model especially tailored to account for the episodic and semantic memory processes underlying recognition responses, in contrast to the 1HTSM (see Batchelder & Riefer, 1990). For example, the occurrence of false recollection, which is precluded by the 1HTSM/2HTSM, is a well-documented phenomenon in memory studies involving lists of semantically-related stimuli (Brainerd & Reyna, 2005; Gallo, 2006). It is possible that the theoretical importance of the overdistribution phenomenon depends on the particular memory task and stimuli being discussed, although even in the case of semantically-related lists the role (even if auxiliary) of a guessing-based account should not be overlooked (e.g., Miller, Guerin, & Wolford, 2011).

Acknowledgments

We thank Charles J. Brainerd for providing his raw data. We also thank Charles J. Brainerd and Bill Batchelder for several insightful comments and suggestions. Model fitting routines used in this manuscript can be obtained upon request.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2014.07.001.

References

Arnold, N. R., Bayen, U. T., Kuhlman, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin and Review, 20*, 326–333.

Bamber, D., & Van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology, 44*, 20–40.

Barchard, K. A. (2012). Examining the reliability of interval level data using root mean square differences and concordance correlation coefficients. *Psychological Methods, 17*, 294–308.

Batchelder, W. H., & Batchelder, E. (2008). Metacognitive guessing strategies in source monitoring. In J. Dunlonsky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 211–244). New York: Psychology Press.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548–564.

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

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.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass: MIT Press.

Brainerd, C. J., Holliday, R. E., Nakamura, K., & Reyna, V. F. (in press). Conjunction illusions and conjunction fallacies in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.

Brainerd, C. J., & Reyna, V. F. (2008). Episodic over-distribution: A signature effect of recollection without familiarity. *Journal of Memory and Language, 58*, 765–786.

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.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106*, 160–179.

Brainerd, C. J., Wang, Z., & Reyna, V. F. (2013). Superposition of episodic memories: Overdistribution and quantum models. *Topics in Cognitive Science, 5*, 773–799.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 307–327.

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

Buchner, A., & Erdfelder, E. (1996). On assumptions of, relations between, and evaluations of some process dissociation measurement models. *Consciousness and Cognition, 5*, 581–594.

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.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer.

Busemeyer, J., & Trueblood, J. (2010). Quantum model for conjoint recognition. In *AAAI fall symposium series. Arlington, VA*. Association for the Advancement of Artificial Intelligence.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge, UK: Cambridge University Press.

Ceci, S. J., Fitneva, S. A., & Williams, W. M. (2010). Representational constraints on the development of memory and metamemory: A developmental-representational theory. *Psychological Review, 117*, 464–495.

Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *Journal of Mathematical Psychology, 53*, 562–576.

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review, 15*, 692–712.

Cox, J. C., & Dobbins, I. G. (2011). The striking similarities standard, distractor-free, and target-free recognition. *Memory & Cognition, 39*, 925–940.

Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumptions in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 531–547.

Curran, T., & Hintzman, D. L. (1997). Consequences and causes of correlations in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 469–504.

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology, 47*, 292–303.

Dodson, C. S., Bawa, S., & Slotnick, S. D. (2007). Aging, source memory, and misrecollections. *Journal of Experimental Psychology: L earning, Memory, and Cognition, 33*, 169–181.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). WordGen: A tool for word selection and non-word generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments & Computers, 36*, 488–499.

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41*, 369–387.

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review, 12*, 403–408.

Feynman, R. P., Leighton, R. B., & Sands, M. (1965). *The Feynman lectures on physics* (Vol. 3). Reading, Mass: Addison-Wesely.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York: Psychology Press.

Gallo, D. L., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51*, 473–493.
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman and Hall.
Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
Grünwald, P. (2007). *The minimum description length principle*. Cambridge, Mass: Mas.
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.
Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*, 185–207.
Hothorn, T., Hornik, K., van de Wiel, M. A., & Zeileis, A. (2008). Implementing a class of permutation tests: The coin package. *Journal of Statistical Software*, 28.
Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.
Jacoby, L. L., & Shrout, P. E. (1997). Toward a psychometric analysis of violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 505–510.
Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12*, 352–361.
Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.
Jones, G. V. (1987). Independence and exclusivity among psychological processes: Implications for the structure of recall. *Psychological Review, 94*, 229–235.
Kellen, D., & Klauer, K. C. (2011). Evaluating models of recognition memory using first- and second-choice responses. *Journal of Mathematical Psychology, 55*, 251–266.
Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin & Review, 20*, 693–719.
Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71*, 7–31.
Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika, 75*, 70–98.
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.
Klauer, K. C., & Kellen, D. (2011). The flexibility of models of recognition memory: An analysis by the minimum-description length principle. *Journal of Mathematical Psychology, 55*, 430–450.
Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin, 26*, 1074–1093.
Lahl, O., Göritz, A. S., Pietrowsky, R., & Rosenberg, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2654 German nouns. *Behavior Research Methods, 41*, 13–19.
Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian data analysis for cognitive science: A practical course*. New York: Cambridge University Press.
Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*, 335–384.
Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1137–1151.
Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E. J. (2013). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*.
Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition, 21*, 121–156.
Meiser, T. (2005). A hierarchy of multinomial models for multidimensional source monitoring. *Methodology, 1*, 2–17.
Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 116–137.
Miller, M. B., Guerin, S. A., & Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1228–1235.
Myung, J. I., Navarro, D. J., & Pitt, M. A. (2006). Model selection by normalized maximum likelihood. *Journal of Mathematical Psychology, 50*, 167–179.
Navarro, D. J. (2004). A note on the applied use of MDL approximations. *Neural Computation, 16*, 1763–1768.
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.
Oravecz, Z., Faust, K., & Batchelder, W. H. (in press). An extended cultural consensus theory model to account for cognitive processes for decision making in social surveys. *Sociological Methodology*.
Peirce, J. W. (2007). PsychoPy-psychophysics toolbox in Python. *Journal of Neuroscience Methods, 162*, 8–13.
Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1110–1115.
Regenwetter, M., & Davis-Stober, C. P. (2012). Behavioral variability of choices versus structural inconsistency of preferences. *Psychological Review, 119*, 408–416.
Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*, 318–339.
Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 680–693.
Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28*, 907–922.
Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104*, 406–415.
Rouder, J. N., Morey, R. D., & Pratte, M. S. (in press). Hierarchical Bayesian models. In W. H. Batchelder, H. Colonius, E. Dzhafarov, & J. I. Myung (Eds.), New handbook of mathematical psychology, Measurement and methodology, Vol. 1. London: Cambridge University Press.
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.
Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General, 137*, 370–389.
Schütz, J., & Bröder, A. (2011). Signal detection and threshold models of source memory. *Experimental Psychology, 58*, 293–311.
Schweickert, R., & Chen, S. (2008). Tree inference with factors selectively influencing processes in a processing tree. *Journal of Mathematical Psychology, 52*, 158–183.
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.
Self, S. G., & Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82*, 605–610.
Singmann, H., Kellen, D., & Klauer, K. C. (2013). Investigating the other-race effect of Germans towards Turks and Arabs using multinomial processing tree models. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1330–1336). Cognitive Science Society.
Sloman, S., Rottenstreich, Y., Wisniewski, E., & Hadjichristidis, C. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 573–582.
Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition, 33*, 151–170.
Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review, 15*, 713–731.
Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54*, 167–183.
Spaniol, J., & Bayen, U. T. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 631–651.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology), 64*, 583–639.
Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 570–586.

Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition, 36*, 1–8.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33*, 203–217.

Thomas, E. A., & Legge, D. (1970). Probability matching as a basis for detection and recognition decisions. *Psychological Review, 77*, 65–72.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science, 5*, 672–688.

Wei, G., & Tanner, M. (1990). A monte carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association, 85*, 699–704.

Wu, H., Myung, J. I., & Batchelder, W. H. (2010a). Minimum description length model selection of multinomial processing tree models. *Psychonomic Bulletin & Review, 17*, 275–286.

Wu, H., Myung, J. I., & Batchelder, W. H. (2010b). On the minimum description length complexity of multinomial processing tree models. *Journal of Mathematical Psychology, 54*, 291–303.

Yu, J., & Bellezza, F. S. (2000). Process dissociation as source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1518–1533.