Journal of Memory and Language 60 (2009) 180–193

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Measuring phantom recollection in the simplified conjoint recognition paradigm

Christoph Stahl *, Karl Christoph Klauer

University of Freiburg, Institute for Psychology, D-79085 Freiburg, Germany

ARTICLE INFO

Article history:
Received 22 January 2008
Revision received 11 August 2008
Available online 30 September 2008

Keywords:
Memory
False memory
Phantom recollection
Conjoint recognition
Multinomial model
Parameter heterogeneity

ABSTRACT

False memories are sometimes strong enough to elicit recollective experiences. This phenomenon has been termed Phantom Recollection (PR). The Conjoint Recognition (CR) paradigm has been used to empirically separate PR from other memory processes. Recently, a simplification of the CR procedure has been proposed. We herein extend the simplified CR paradigm to the measurement of PR by including an additional parameter. Two experiments in the simplified CR paradigm were conducted in which participants studied lists of items that converge on a single semantic associate. PR was obtained for lists from which eight items were presented but not for lists from which a single item was presented. In addition, new evidence for the validity of the guessing processes in the simplified CR paradigm is reported. These findings support the validity and usefulness of the simplified CR model as a measurement tool for processes of veridical and false memory.

Introduction

Human observers can sometimes retrieve the meaning or gist of a statement without being able to recollect any surface detail. This is a core assumption of Fuzzy Trace Theory (FTT; e.g., Brainerd & Reyna, 2002). According to FTT, separate memory traces for the meaning or *gist* of an item (e.g., its semantic category) and its identity or *verbatim* detail (e.g., its exact wording) can be retrieved independently from memory. The distinction between gist and verbatim traces has proven to be important in research on false memory (e.g., Brainerd, Forrest, Karibian, & Reyna, 2006; Brainerd, Payne, Wright, & Reyna, 2003; Brainerd & Reyna, 2002; Brainerd & Wright, 2005; Brainerd, Wright, Reyna, & Mojardin, 2001; Odegard & Lampinen, 2005; Seamon et al., 2002; Wright & Loftus, 1998). For instance, FTT can accommodate a phenomenon in semantic false recognition that has been termed *phantom recollection* (Brainerd et al., 2001). Phantom recollection describes the phenomenon that, when gist-based false memories arise at high levels, a subset of those false memories may be accompanied by false recollective experiences, as has often been reported in research using the Deese–Roediger–McDermott paradigm (DRM; Deese, 1959; Roediger & McDermott, 1995). A distracter that is semantically related to an old item (i.e., a target) and elicits such vivid false memories is likely to be mistaken as the target.

To separate verbatim and gist memory empirically, the Conjoint Recognition (CR) paradigm and multinomial model have been proposed (Brainerd, Reyna, & Mojardin, 1999; Brainerd, Stein, & Reyna, 1998). We recently introduced a simplified and considerably more efficient version of the original CR paradigm that allows for the separation of verbatim and gist memory in a single group of participants (as opposed to three groups of participants required in the original CR paradigm). The simplified CR paradigm has been shown to provide valid parameter estimates for verbatim and gist memory for targets, gist-based acceptance of related distracters, and recollection rejection of related distracters, as well as parameters for two relevant guessing processes (Stahl & Klauer, 2008; see also Brainerd, Reyna, Bellinge, & Myers, 2008, for a related simplification of the CR paradigm using a repeated-measures approach). Our

* Corresponding author. Fax: +49 761 203 2417.
E-mail address: stahl@psychologie.uni-freiburg.de (C. Stahl).

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.08.001

initial model could not accommodate phantom recollection. In the present work, an extended version of the model is introduced that includes a parameter for phantom recollection, and we report the results of two experiments in the simplified CR paradigm that demonstrate the validity of the new parameter. In addition, further evidence for the validity of the guessing processes is presented.

The CR paradigm and model

In Brainerd et al.'s (1999) original CR paradigm, participants are first presented with a study list. The test list contains three types of items: Target probes (i.e., old items from the study list), *related distracters* that share a target's gist, and *unrelated distracters*. The memory test is administered to three groups of participants with different instructions: Under the T instruction, participants are asked to accept only targets; under the R instruction, participants are to accept as old only related distracters; and, finally, under the T+R instruction, participants are to accept both targets and related distracters. From the proportions of accepted targets, related distracters, and unrelated distracters obtained in the three different groups, the parameters of a multinomial model are estimated that provide measures of verbatim and gist memory.

Recently, we introduced a simplified CR paradigm and model that also provides valid estimates of verbatim and gist memory, but on the basis of a much simpler procedure (Stahl & Klauer, 2008). In the simplified paradigm, as in the original paradigm, participants are presented with targets, related probes, and unrelated probes, and they are informed as to the types of items that compose the test list. However, in contrast to the original paradigm, the simplified paradigm does not require acceptance or rejection responses, but the identification of the probe's type. Participants are asked to classify probe items as either targets, related probes, or unrelated probes. They are instructed to respond 'target' if they believe that the current probe has been presented in the learning phase. If they believe the current probe to be a related distracter, they are instructed to indicate this by selecting the 'related' response. If they consider the probe to be an unrelated distracter, they are to select the 'new' response. In the simplified CR paradigm, verbatim and gist memory can be separated using the data from only one group of participants. It is therefore considerably more efficient than the original CR paradigm in which three groups of participants are required.

The simplified CR model is depicted in Fig. 1. Consider the first tree diagram that represents the cognitive processes occurring when a target probe is presented at test. In case of available verbatim memory, a target is correctly identified as such. Given no verbatim memory but available gist memory—participants have identified the probe's meaning as old but cannot remember whether the probe itself or a related item with the same gist had been pre-

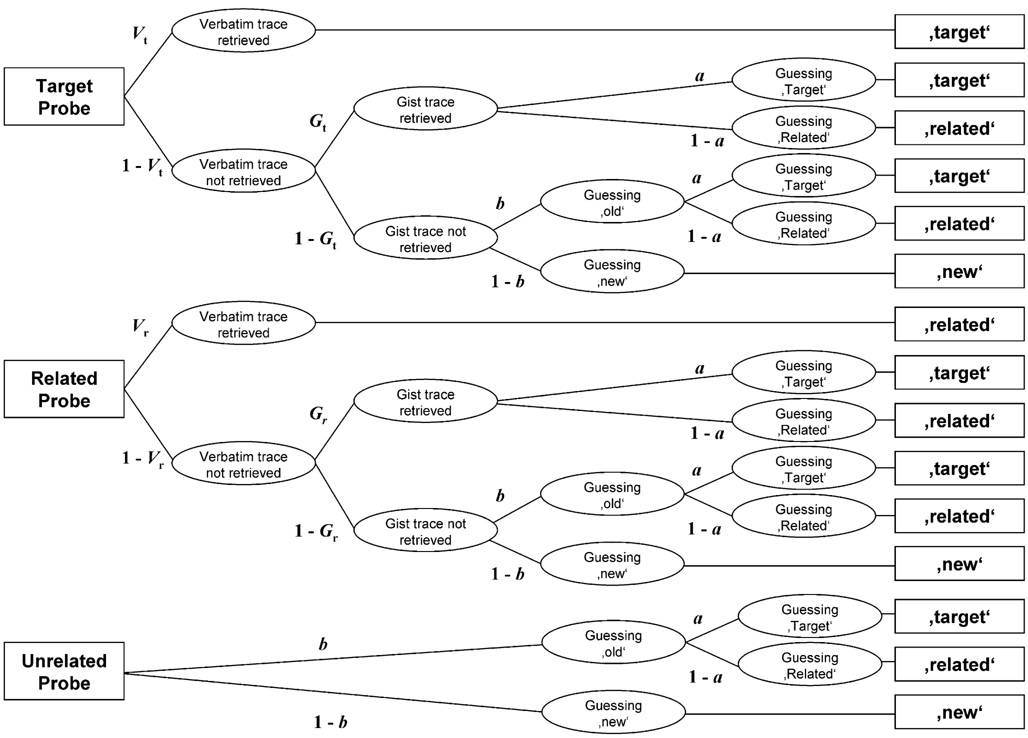

Fig. 1. Processing tree model for the simplified CR paradigm. Rectangles on the left denote probe type, rectangles on the right denote responses. They are connected by branches of the processing tree that represent the combination of cognitive processes postulated by the model. V_t = probability of retrieving a target's verbatim trace given a target probe; V_r = probability of retrieving a target's verbatim trace given a related probe; G_t = probability of retrieving a target's gist trace given a target probe; G_r = probability of retrieving a target's gist trace given a related probe; b = probability of guessing that an item is either a target or a related probe; a = probability of guessing 'target'.

sented in the learning phase—a decision has to be made between the 'target' and 'related' response options. With probability a, the probe is classified as a target, and with probability $1 - a$, the probe is classified as a related distracter. Should neither verbatim nor gist memory be available, participants can still guess that the item's meaning is old (with probability b). In this case, a choice between the 'target' and 'related' responses is again required and captured by the new parameter a as described above. The same decision process involving parameter a is postulated to occur for related probes, as can be seen in the branches $(1 - V_r)G_r$ and $(1 - V_r)(1 - G_r)b$ of the second diagram. Classifications of unrelated probes are based on a combination of guessing processes a and b as illustrated in the third diagram.

The original CR model has been extended to measure phantom recollection by adding a parameter P_r to the processing tree for related probes (Brainerd et al., 2001). The present simplified CR model can analogously be extended to measure phantom recollection. Fig. 2 illustrates how the parameter for phantom recollection, P_r, is integrated into the processing tree for related probes of the simplified CR model (second tree diagram). In the extended version of the simplified CR model, phantom recollection occurs with probability $(1 - V_r)P_r$, describing the probability of a *target* response to a related distracter in the absence of recollection rejection. Gist-based false memories that lack such vivid recollective experiences as are required for phantom recollection would occur with probability $(1 - V_r)(1 - P_r)G_r$ and would be followed by the guessing process modeled by parameter a. In the absence of memory, related distracters can be accepted as old on the basis of guessing processes with probability $(1 - V_r)(1 - P_r)(1 - G_r)b$, and guessing as modeled by parameter a then determines classification as target versus related probe. Note that the model in Fig. 2 has seven parameters while the data yields only 6 degrees of freedom (i.e., two free empirical probabilities each for targets, related probes, and unrelated probes), implying that the model is not identified without additional degrees of freedom or restrictions on (some of) the parameters. The issue of identifiability is discussed below, and the model is formally specified in Appendix A.

We have demonstrated the applicability of the thus extended version of our CR model by reanalyzing data from Experiment 2 of Stahl and Klauer (2008; see General Discussion). In that experiment, we manipulated gist activation by including in the study list either one (i.e., the Weak Gist activation condition) or four items (i.e., Strong Gist activation) from a given DRM list. Two P_r parameters were added to the simplified CR model, one for each level

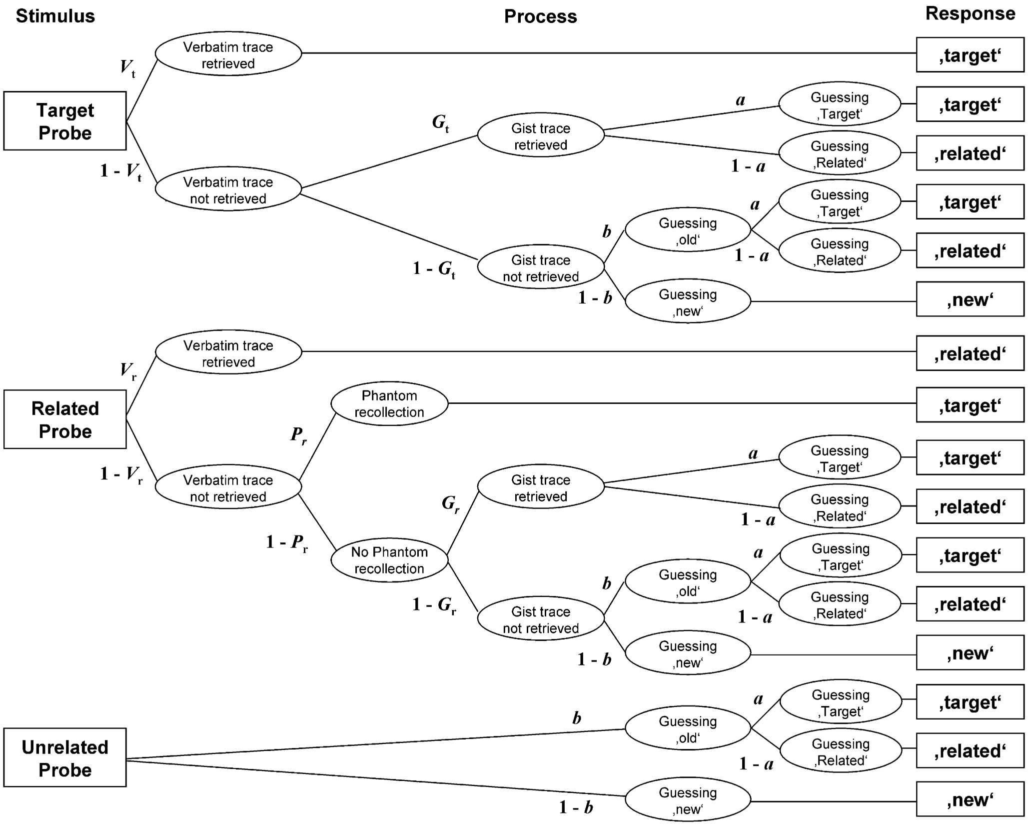

Fig. 2. Extension of the simplified CR model for the measurement of phantom recollection of related probes. Rectangles on the left denote probe type, rectangles on the right denote responses. They are connected by branches of the processing tree that represent the combination of cognitive processes postulated by the model. V_t = probability of retrieving a target's verbatim trace given a target probe; V_r = probability of retrieving a target's verbatim trace given a related probe; P_r = probability of phantom recollection; G_t = probability of retrieving a target's gist trace given a target probe; G_r = probability of retrieving a target's gist trace given a related probe; b = probability of guessing that an item is either a target or a related probe; a = probability of guessing 'target'.

of the Gist Activation factor. Estimates of P_r were at .11 for Weak Gist and at .15 for strongly activated gist. However, P_r parameters were not affected by the number of items, and they were also not significantly different from zero, suggesting that presenting four DRM items did not suffice to obtain substantial phantom recollection. In the present paper, we report the results of two experiments in which we obtained significant phantom recollection as measured by the simplified CR paradigm and model as depicted by Fig. 2 when eight items were presented.

Guessing processes in the simplified CR paradigm

Both the original and the simplified CR paradigms include parameters for guessing processes. In the original CR paradigm, when neither verbatim nor gist memory are retrieved, participants can accept a probe as on the basis of guessing. In this case, participants take the probe for an 'old' item by way of guessing with probability b; with probability $1 - b$, they consider the probe to be a 'new' item.

In the simplified CR paradigm, participants are asked to classify probes as targets, related distracters, or unrelated distracters. In what follows, and departing from common usage, we will refer to related distracters and targets as old because they refer to old meaning, and to unrelated distracters as new. Similar to the original CR paradigm, participants guess that a probe is a new distracter, and thus select the 'new' response by way of guessing, with probability $1 - b$ in the simplified CR paradigm. However, guessing processes differ between the original and the simplified CR models when a probe is judged to be old by way of guessing. In the simplified paradigm, when a probe is taken to be old by way of guessing with probability b or when its gist is recognized, participants have to guess whether it is a target or a related distracter. This guessing process is modeled by parameter a such that participants select the 'target' response with probability a and prefer the 'related' response with probability $1 - a$.

Guessing processes were discussed and validated in terms of a Bayesian metacognitive framework proposed by Batchelder and Batchelder (2008). A formal application of this framework to the simplified CR paradigm is given in Stahl and Klauer (2008). In brief, this framework assumes that participants base their guessing strategies on base rates of different types of probes on the test list, and on the relative strength of their memory for different types of items. In the simplest case, participants try to match their response frequencies to the base rates of the different probes. For example, if there are 50% old and 50% new items on a test, they try to respond 'old' and 'new' in about 50% of cases if and when they guess. This is known as 'probability matching' (e.g., Greene, 1996; see also Batchelder & Batchelder, 2008). In the present work, experimental evidence for the validity of the guessing parameters is provided by manipulating base rates of different probe types at test.

In addition to base rates, participants may also use knowledge about their own memory to inform their guessing strategies (i.e., metacognitive guessing strategies). As outlined above, when a probe is taken to be old because its gist has been recognized, participants have to guess whether it is a target or a related distracter. Given that verbatim traces are likely to be more accessible for targets than for related probes, it follows that a situation in which gist memory is accessible but verbatim memory is not is more likely to occur for related probes than for target probes. Thus, when a probe is judged to be old on the basis of available gist memory, but verbatim memory is lacking, the probe is more likely to be a related distracter than a target. If participants use this metacognitive information, guessing is expected to be biased towards related probes—or, more generally, the class of items for which verbatim memory is weakest. This is what we observed across six experiments in our introduction of the simplified CR paradigm (Stahl & Klauer, 2008).

The present work introduces and validates an extension of the simplified CR model (Stahl & Klauer, 2008) towards measuring phantom recollection, and it provides additional evidence for the validity of the model's guessing parameters. Two experiments were conducted in which phantom recollection was provoked using the DRM methodology. Orthogonally, guessing bias was manipulated by varying the proportion of targets, related probes, and unrelated probes at test. Before we turn to the experiments, two issues need to be addressed: the identifiability of the model, and the role of the recollection-rejection process.

Identifiability

As pointed out above, the model in Fig. 2 has seven parameters—V_t, V_r, G_t, G_r, P_r, a, and b. The data (i.e., 'target', 'related', and 'new' responses to targets, related probes, and unrelated probes) yield only six free empirical probabilities (for each probe type, the response probabilities have to sum up to one). More specifically, there are now three free memory parameters in the tree for related probes that appear only in this tree, but the data obtained from related probes yield only two free empirical probabilities. The model as depicted in Fig. 2 is therefore not identified (i.e., there is not one unique solution for the parameter estimates). This means that the number of free parameters must be reduced by imposing restrictions on some of the parameters in a way that leads to an identifiable version of the model, before the model can be applied to data. This problem stems from the tree for related probes; the trees for target probes and for unrelated probes have not been modified, and the parameters in these trees remain identified as has been shown in Stahl and Klauer (2008).

The number of free parameters can be reduced by fixing one or more parameters to constant values, or by equating two or more parameters. By imposing a-priori restrictions on the parameters of a model, psychological assumptions are made about the processes represented by these parameters. It is important that these assumptions are plausible and well-supported; otherwise, the resulting model will be misspecified and its parameter estimates may be systematically biased. This requirement narrows the set of possible restrictions on the model parameters down to a few psychologically plausible ones.

One plausible assumption is made in the simplified CR paradigm as introduced in Stahl and Klauer (2008) and illustrated in Fig. 1. Note that the tree for related probes shown in Fig. 1 can be considered a special case of the respective tree shown in Fig. 2, with the additional restriction that the P_r parameter is fixed to zero. In restricting $P_r = 0$, the assumption is made that phantom recollection does not figure in a typical study using the simplified CR paradigm. This assumption has received support across a series of validation studies (Stahl & Klauer, 2008).

Another plausible assumption is that the recollection-rejection process does not contribute to performance in the present experiments. This would allow for the restriction $V_r = 0$ (and would render the model in Fig. 2 identifiable). In the present study, this restriction had the status of an auxiliary assumption that was not necessary for the identifiability of the model but was nevertheless introduced to reduce the number of free parameters in the tree for related probes and thereby increase the precision with which phantom recollection was measured. We will present support for the psychological plausibility of this assumption in the next section. Empirical tests of the assumption are reported in the Results sections.[1]

In the present study, identifiability was ensured by equating memory parameters across the levels of a base rate manipulation targeted at the guessing parameters (formal proof for the model's identifiability is given in Appendix A). It was thereby assumed that a base-rate manipulation implemented at test did not affect participants' memory processes, an assumption that is supported by a large body of research in which base-rate effects were found only in response bias but not in memory performance (e.g., Estes & Maddox, 1995; Healy & Kubovy, 1978; Hirshman & Henzler, 1998; Rhodes & Jacoby, 2007).

In addition, the assumption was made that the same parameters govern guessing processes for probes from DRM lists from which one item had been presented, and for probes from DRM lists from which eight items had been presented. Support for these assumptions comes from a series of validation studies in which we have shown that the parameters of the simplified CR model are valid measures of the processes they represent (Stahl & Klauer, 2008), and this ensemble of assumptions is empirically tested for the present data by means of the model's goodness-of-fit test reported below.

Recollection rejection

An auxiliary assumption made in the present applications of the model is that the contribution of the recollection-rejection process to the present data is negligible. This assumption is based on both empirical and theoretical considerations. Empirically, our own previous research has shown that recollection rejection, as measured by parameter V_r, did not contribute substantially to performance in a paradigm such as the ones implemented here. In our set of validation studies of the simplified CR paradigm, V_r was significantly different from zero only when a priming manipulation was used that was explicitly targeted at increasing levels of recollection rejection (Stahl & Klauer, 2008). Further support comes from other researchers' findings, suggesting that recollection rejection is even less likely to occur when DRM lists are used (Gallo, 2004). Given that recollection-rejection is not found in typical studies in the simplified CR paradigm (i.e., with moderate to low levels of false memory), it is difficult to see why one would expect substantial recollection-rejection in experimental conditions promoting high levels of false memory (but see Brainerd et al., 2001, for diverging results obtained with the original CR paradigm).

Theoretically, the recollection-rejection process is not well-specified for a situation in which a related probe is associated with more than one target. Recollection-rejection is thought to occur when a related lure (e.g., *ROSE*) is presented at test, and its gist cues the retrieval of the corresponding target's verbatim trace (e.g., *TULIP*). In this case, participants can reject the related lure because, given their knowledge of the target, the lure cannot possibly have been presented. But this characterization of the recollection-rejection process runs into problems when, as with DRM lists, multiple targets are presented: which one of the targets is associated with the related probe, and is to be retrieved? Gallo (2004) argued that recollection-rejection of critical lures from DRM lists should occur only when participants can exhaustively recall all of the presented list items (e.g., they can recollect all of the flower names presented at study), because only then it is clear that the probe cannot have been presented in the study phase (i.e., it disqualifies as a potential target).

Gallo (2004) has shown that, as predicted from these considerations, such recollection-rejection processes are negligible when four target items were presented from a given DRM list. Given that the present studies presented twice as many items from each DRM list, thereby decreasing the likelihood of exhaustive recall even further, it is implausible to assume that substantial recollection rejection should occur in the present studies. In sum, previous studies suggest that V_r should be zero in a typical study in the simplified CR paradigm, and there is both empirical and theoretical evidence supporting the assumption that V_r should be even smaller (or, at the very least, not greater) when multiple targets are associated with a related probe. We therefore felt that it was safe to make the assumption that recollection-rejection processes should not figure in accounting for the present data. As a first step in the model analyses reported below, this assumption was empirically evaluated for the present data by way of a statistical comparison of an unrestricted model's goodness-of-fit of with the fit of a model in which V_r was restricted to zero. The assumption proved to be justified in both data sets.

Experiment 1

In Experiment 1, we aimed to validate the new phantom recollection parameter P_r. We implemented a manipulation that has been shown to affect phantom recollection in the original CR paradigm. Brainerd et al.

[1] A second set of analyses was also performed with a model in which P_r was fixed to zero for the Weak-Gist condition and the V_r parameter was equated across the Gist Activation factor. This set of analyses yielded the same pattern of significant and nonsignificant results; additional detail can be obtained from the first author.

(2001) demonstrated phantom recollection using DRM lists: When the theme of a DRM list was repeatedly activated at study by multiple list items, phantom recollection for the critical lure was obtained in addition to gist-based false memory as measured by parameter G_r. In the present Experiment 1, phantom recollection was manipulated by varying the number of items (one vs. eight) that were presented at study from a given DRM list. Phantom recollection was predicted to be evident for critical lures from DRM lists from which eight items were presented at study but not for critical lures from DRM lists from which only one item was presented. Gist-based false memory (parameter G_r) is predicted to be greater when eight rather than one item are presented from a DRM list.

DRM lists consist of semantic associates that converge on critical, nonstudied lure words (e.g., butter, loaf, knife—bread), and false memory for these critical lures (e.g., bread) is often as strong as veridical memory for presented items (e.g., butter). Importantly, the logic behind the present studies depends on the assumption that critical lures are strong semantic associates of list items, but strong semantic associations do not exist among list items. Note that there may also be some degree of semantic associations between list items, and thus, the DRM procedure may also activate the gist of some of the list items that are used as target probes to some degree; however, these effects are usually weaker and less consistent than those for the critical lures. For the materials used here, we have shown that such effects are negligible (Stahl & Klauer, 2008, Exp. 2). In sum, presenting an increasing number of DRM list items strengthens the activation of the critical lure's gist but not of the gist of other list items. In terms of the parameters of the present model, increasing the number of presented DRM list items should affect parameters for gist-based false memory, such as G_r and P_r, but should not affect gist-based veridical memory as measured by parameter G_t.

We also manipulated the proportion of related and unrelated probes on the test list as a between-subjects factor. In the high a/low b condition, the ratio of targets to related distracters was 1:1, and the ratio of "old" probes (i.e., targets and related distracters) to "new" probes (i.e., unrelated distracters) was 1:2. In the low a/high b condition, the ratio of targets to related probes was reduced to 1:4, and the ratio of "old" to "new" probes was increased to 5:1. It is predicted by the Bayesian framework that participants' guessing strategies will be informed by the base rates of the different types of items. Thus, we expect that participants will be more likely to guess 'target' when the ratio of targets to related probes is high than when it is low; and participants will be less likely to guess 'old' when the ratio of old to new probes is low than when it is high. We therefore predict higher values of a, and lower values of b, in the high a/low b condition as compared to the low a/high b condition.

Method

Participants

Eighteen volunteers participated (10 females; ages 20–34, $M = 25$). Participants were sampled from the department's database of volunteers (mostly students from Freiburg's universities and colleges, as well as non-student citizens) and participated in exchange for a certificate of participation or a monetary compensation of 3.50 Euro. Participants' native language was German.

Design

A 2 (Base Rate: high a/low b vs. low a/high b) $\times$ 2 (Gist Activation: weak vs. strong) $\times$ 3 (Probe Type: target, related, unrelated) design was implemented with repeated measures on the last two factors.

Materials

We used German DRM lists (Deese, 1959; Roediger & McDermott, 1995) that were taken from Stegt (2006). DRM lists consist of a number of to-be-presented items and a single critical lure item that is related to list items by common gist. The critical lure items are not presented but nevertheless often recalled and recognized with high probability. Thirty-six DRM lists were used, each consisting of 11 list items and one critical lure. The first list item of each presented list was used as a target probe, and the critical lure as a related distracter; the first list items of non-presented lists were used as unrelated distracters.

The study list presented items from 24 randomly selected DRM lists that were randomly split into two halves of 12 DRM lists each. Gist Activation was manipulated by presenting different numbers of items from each DRM list at study. The first half of the DRM lists was used for the Weak condition in which a DRM list was represented by a single item; the second half was used for the Strong condition in which a DRM list was represented by eight items. DRM lists from the first half were represented by the first DRM list item, DRM lists from the second half by the first eight items. In total, 108 items were thereby presented. Order of lists was randomized. Five items were added as primacy buffer and five items as recency buffer to the study lists.

At test, the 12 single-item DRM lists and the 12 eight-item DRM lists were each randomly split into 6 DRM lists for which the first DRM list item was used as target probe and 6 DRM lists for which the critical lure was used as related probe. Unrelated probes were the first DRM list items from 12 DRM lists that were not represented in the study list.

To manipulate the proportion of unrelated probes, three additional list items from each non-presented list were used as unrelated probes in the high a/low b group. In sum, 36 additional unrelated probes were presented in the high a/low b group, thereby increasing the base rate of unrelated probes, which is predicted to increase the probability $1 - b$ (i.e., decrease the probability b) with which participants respond "new" by way of guessing processes.

To manipulate the ratio of target probes to related probes on the test list, three additional list items were probed from each of the presented DRM lists in the low a/high b group. These related probes were not critical lures but list items that were not presented on the study list. In sum, 36 additional related probes were presented in the low a/high b condition. Data from these additional items were not analyzed. In total, in each group, 72 items were thereby presented at test in random order. All randomizations were carried out for each participant anew.

Procedure

Experiment 1 implemented the simplified CR procedure (Stahl & Klauer, 2008). Participants were sampled in individual computerized sessions. They were told that they would see a list of items that they had to remember for a later test. Study items were presented sequentially for 4000 ms in black sans-serif letters on a grey background in the center of the screen. DRM lists were presented in a forward order, such that the first item from each DRM list that typically has the strongest association with the critical lure was presented first. Between DRM lists, a string of asterisks was presented for 4000 ms that indicated that presentation of a new list was about to begin. After the study phase, participants worked on simple arithmetic problems for a total duration of five minutes, after which the memory test was administered. Participants were informed about the three types of probes that composed the test list, and about the proportions with which these probes occurred on the test list. In the memory test, participants were presented sequentially with a list of probes and asked to indicate whether the probe was identical to an old item (i.e., a 'target'), 'related' to an old item, or 'new'. Participants classified the probes by selecting the appropriate response (i.e., 'target', 'related', or 'new') with a computer mouse. After completing the memory test, participants were thanked, debriefed, and dismissed.

Results

Parameter estimation and hypotheses tests reported below were performed with the HMMTree software (Stahl & Klauer, 2007). Sensitivity power analyses (performed with G∗Power 3; Faul, Erdfelder, Lang, & Buchner, 2007) assured high test power, $1-\beta = .95$, for parameter comparisons across conditions. With $\alpha = \beta = .05$, we were able to detect small to medium effects ($w = .17$; see Cohen, 1988, Chap. 5).

The model depicted in Fig. 2 has seven parameters—five memory parameters (V_t, G_t, V_r, G_r, and P_r) and two guessing parameters (a and b). It was adapted to the present study by allowing the parameters to vary across the cells of the 2 (Base Rate; manipulated between-subjects) × 2 (Gist Activation; manipulated within-subjects) design as follows. First, the memory parameters were allowed to vary across the two levels of the Gist Activation factor (yielding five additional parameters); the memory parameters were set equal across the Base Rate factor because the base rate manipulation should not affect memory processes.

Second, guessing parameters were allowed to vary across the two levels of the Base Rate factor (yielding two additional guessing parameters). Guessing parameters were set equal across the Gist Activation factor because guessing parameters should be relevant only in the absence of information in memory. In total, the model had 14 parameters (i.e., the 7 parameters of the extended CR model, plus 5 additional memory parameters to accommodate the effects of the Gist Activation manipulation, plus 2 additional guessing parameters to accommodate the effects of the Base Rate manipulation).

The experimental design yielded 20 free empirical probabilities (i.e., two free empirical probabilities for each of the five probe types—target probes from 1-item and 8-item lists, critical lures from 1-item and 8-item lists, and unrelated probes—in each of the two conditions). Given 14 parameters, there are 6 *df* for a test of the model's goodness of fit.

The model, including the above restrictions, fitted the data well, $G^2_{(df=6)} = 7.14$, $p = .31$. Goodness-of-fit was slightly improved in terms of the p value when V_r was set to zero, $G^2_{(df=8)} = 8.69$, $p = .37$; the assumption of negligible recollection-rejection proved to be justified, $\Delta G^2_{(df=2)} = 1.55$, $p = .46$. Parameter estimates and significance tests are given in Table 1. As predicted, phantom recollection (parameter P_r) was greater for critical lures from DRM lists from which eight list items had been presented at study (i.e., the Strong Gist condition) than for critical lures from DRM lists from which only a single item had been presented (i.e., the Weak Gist condition). Phantom recollection was not significantly different from zero in the Weak Gist condition, $\Delta G^2_{(df=1)} < 0.01$, $p > .92$, but it was significantly different from zero in the Strong Gist condition, $\Delta G^2_{(df=1)} = 10.78$, $p < .01$. This finding also implies that the simpler model depicted in Fig. 1 without a parameter for phantom recollection did not provide an adequate fit to the data, $G^2_{(df=10)} = 19.47$, $p < .05$.

Furthermore, also as predicted, guessing was affected by the base rate manipulation. Parameter a was greater in the high a/low b condition than in the low a/high b condition, $\Delta G^2_{(df=1)} = 5.41$, $p < .05$. Conversely, parameter b was greater in the low a/high b condition than in the high a/low b condition, $\Delta G^2_{(df=1)} = 6.63$, $p < .05$.

In addition to phantom recollection, gist-based false memory (parameter G_r) was also strongly and significantly increased in the Strong Gist condition. Gist memory parameter G_t was not affected by the Gist Activation factor. Unexpectedly, verbatim memory for targets (parameter V_t) was also significantly affected by the Gist Activation factor, $\Delta G^2_{(df=1)} = 4.93$, $p < .05$.

Discussion

The results of Experiment 1 largely confirm our predictions and provide support for the validity of the phantom recollection parameter P_r. We obtained substantial phan-

Table 1
Estimates (and 95% confidence intervals) of the parameters of the simplified CR model for Experiment 1

Parameter	Gist activation		$\Delta G^2_{(df=1)}$	p
	Weak	Strong		
$a_{\text{high } a/\text{low } b}$	.26 (.13, .39)			
$b_{\text{high } a/\text{low } b}$	.22 (.14, .29)			
$a_{\text{low } a/\text{high } b}$	.11 (.02, .20)			
$b_{\text{low } a/\text{high } b}$	.35 (.27, .43)			
G_t	.47 (.25, .69)	.35 (.07, .64)	0.44	.51
G_r	.12 (.00, .27)	.80 (.70, .91)	47.55	<.01
P_r	.00 (.00, .06)	.21 (.08, .34)	10.78	<.01
V_t	.60 (.49, .71)	.75 (.67, .84)	4.84	.03

Note. V_t = probability of retrieving a target's verbatim trace given a target probe; G_t = probability of retrieving a target's gist trace given a target probe; G_r = probability of retrieving a target's gist trace given a related probe; b = probability of guessing that an item is either a target or a related probe; a = probability of guessing 'target', P_r = probability of classifying a related probe as a target based on phantom recollection.

tom recollection of critical lures when eight items from a DRM list were presented but not when a single item was presented from a DRM list. Gist-based false memory as measured by parameter G_r was also affected by the gist manipulation in the predicted manner.

The results replicate an important finding reported by Stahl and Klauer (2008; Experiment 2), who report significant effects on G_r but not on G_t when DRM lists were used to manipulate gist memory. Note that the DRM lists are created such that the gist of the list items converges on the gist of the critical lure. In other words, associations between list items and the critical lure are maximized, such that the presentation of each list item activates the critical lure's gist. In contrast, associations among the list items themselves are not maximized and are generally much weaker. This is reflected in the pattern of estimates for the G_r and G_t parameters: Parameter G_r, reflecting activation of the critical lure's gist, is strongly increased when seven additional DRM list items were presented; in contrast, parameter G_t, reflecting activation of first list item's gist, is not significantly affected by the presentation of seven additional list items.

The results also provide support for the validity of the guessing parameters. Both guessing parameters were affected in the predicted direction by the manipulation of the relevant base rates. In addition, the effects on the guessing parameters were opposite in direction, suggesting that they measure independent and dissociable guessing processes. The predictions of the Bayesian metacognitive framework were thus confirmed for parameter a: It was demonstrated that parameter a is sensitive to the ratio of targets to related distracters on the test list. However, a second possibility should be considered: Although unlikely, it might be argued that the effect on parameter a was not due to the manipulation of the ratio of targets to related probes but to the manipulation of the ratio of old to new probes. We will address this possibility in Experiment 2.[2]

[2] As suggested by one reviewer, it is possible that the guessing process modeled by parameter a was affected by the Gist Activation manipulation. That is, if participants were aware for a given probe how many related items were presented at study (one vs. eight), this knowledge might affect their tendency to select the 'target' versus 'related' responses. That is, it is more likely that a probe is a target when it is associated with an eight-item list, simply because there were eight possible targets, as compared with one target for one-item lists. In terms of processes, guessing 'target' versus 'related' (parameter a) might be increased for probes from eight-item lists when the probe's gist was retrieved together with some information about the length of the list that is associated with this gist (i.e., a very strong experience of familiarity, or retrieval of at least two targets from that list). To address this concern empirically, we fitted models with different a-parameters for the strong-gist items. In these models, V_r was fixed to zero, and P_r was fixed to zero for the Weak-Gist condition only. In a first model, a separate a-parameter was introduced for the Strong-Gist condition. The separate guessing parameters for the Strong-Gist condition did not differ from the parameters of the Weak-Gist condition, $\Delta G^2(2) = 0.19$ and 3.31, for Experiments 1 and 2, respectively, both $ps > .19$. In a second model, a separate a-parameter was introduced only for the guessing process following gist retrieval (i.e., the G_r branch) but not for the guessing process when gist was not retrieved (i.e., the $1 - G_r$ branch). The separate guessing parameters could be set equal to those in the Weak-Gist condition, $\Delta G^2(2) = 0.19$ and 3.35, for Experiments 1 and 2, respectively, both $ps > .19$. Thus, we felt it was safe to conclude that there was no difference in the guessing process modeled by parameter a between the Weak- and the Strong-Gist conditions.

An unexpected increase in verbatim memory was observed in the Strong Gist condition. This effect is relevant because it is a potential threat to the validity of the model. Specifically, if a gist manipulation as used here affects parameter V_t, this parameter cannot be considered a valid measure of verbatim memory. However, the validity of the V_t parameter has been demonstrated (Stahl & Klauer, 2008). We believe that the effect on V_t was due to our specific procedure and reflected an actual increase in verbatim memory in the Strong Gist condition. In the study phase of Experiment 1, participants were told that they would read a number of different word lists, and that they should memorize them for a later test. They were then presented with DRM lists that were represented by either one or eight items, with all items from a list being presented in direct sequence. Importantly, the beginning of a new list was indicated by a row of asterisks. In our view, marking the beginning of a new list increased the salience of the first list item and induced rehearsal strategies that lead to the verbatim memory effect. More precisely, because the first items from each list (which were also used as target probes) were made especially salient, participants tended to rehearse those items more often than subsequent list items. They tended to rehearse items from a given list until a new list was presented. Because presentation of eight-item lists took longer than presentation of one-item lists, participants had more time to rehearse the first items from eight-item lists, and their verbatim memory was better for these items as compared to items from one-item lists.

Experiment 2

We tested this assumption in a second experiment. If the effect on V_t is due to the fact that participants were told when presentation of a new list would begin, removing this information should eliminate the V_t effect. In Experiment 2, all items were presented in a single sequence uninterrupted by asterisks.

An additional modification concerned the base rate manipulation. In order to test the possible alternative explanation of the effect on parameter a, we held the ratio of targets to related probes constant across conditions while manipulating only the ratio of old to new probes. If the predictions derived from the Bayesian framework hold true, parameter a should not be affected; if, instead, parameter a is sensitive to the ratio of old to new probes, the same effect on parameter a as in Experiment 1 should be obtained. As in Experiment 1, parameter b should be sensitive to the ratio of old to new probes. We therefore predicted a greater estimate of b in the high b as compared to the low b condition.

Method

Participants

Twenty new volunteers participated (11 female; ages 18 to 34, $M = 22$) in exchange for a certificate of participation or a monetary compensation of 3.50 Euro. Participants' native language was German.

Design

A 2 (Base Rate: low b vs. high b) × 2 (Gist Activation: weak vs. strong) × 3 (Probe Type: target, related, unrelated) design was implemented with repeated measures on the last two factors.

Materials

The same materials as in Experiment 1 were used with the following changes: In Experiment 2, the proportion of related probes on the test list was not manipulated, thereby holding constant the ratio of target probes to related probes. Therefore, the test list always contained 12 related probes (6 for each level of the Gist Activation factor). As a consequence of our manipulation of the proportion of unrelated probes, this results in a total of 36 probes in the high b condition and 72 probes in the low b condition.

Procedure

Procedure was identical to that of Experiment 1, with the exception that asterisks indicating the beginning of a new list were not presented in Experiment 2.

Results

Memory parameters were set equal across the Base Rate factor, and guessing parameters were set equal across the Gist Activation factor. Model fit was very good, both for the unrestricted model, $G^2_{(df=6)} = 3.65$, $p = .72$, and for the restricted model, $G^2_{(df=8)} = 4.57$, $p = .71$; restricting $V_r = 0$ again proved to be justified, $\Delta G^2_{(df=2)} = 0.92$, $p = .63$. Parameter estimates and significance tests are given in Table 2. Replicating Experiment 1, phantom recollection (parameter P_r) was again greater in the Strong than in the Weak Gist condition. Phantom recollection was not significantly different from zero in the Weak Gist condition, $\Delta G^2_{(df=1)} = 0.38$, $p = .54$, but it was significantly different from zero in the Strong Gist condition, $\Delta G^2_{(df=1)} = 17.98$, $p < .01$. This finding again implies that the simpler model depicted in Fig. 1 without a parameter for phantom recollection did not provide an adequate fit to the data, $G^2_{(df=10)} = 22.93$, $p < .05$.

Table 2
Estimates (and 95% confidence intervals) of the parameters of the simplified CR model for Experiment 2

Parameter	Gist activation		$\Delta G^2_{(df=1)}$	P
	Weak	Strong		
$a_{\text{low } b}$	.29 (.15, .42)			
$b_{\text{low } b}$	.25 (.18, .32)			
$a_{\text{high } b}$	.24 (.14, .35)			
$b_{\text{high } b}$	.41 (.33, .49)			
G_t	.35 (.13, .56)	.44 (.22, .67)	0.41	.52
G_r	.09 (.00, .26)	.65 (.50, .80)	25.22	<.01
P_r	.03 (.00, .11)	.33 (.20, .47)	16.32	<.01
V_t	.50 (.39, .61)	.60 (.49, .71)	1.73	.19

Note. V_t = probability of retrieving a target's verbatim trace given a target probe; V_r = probability of retrieving a target's verbatim trace given a related probe; G_t = probability of retrieving a target's gist trace given a target probe; G_r = probability of retrieving a target's gist trace given a related probe; b = probability of guessing that an item is either a target or a related probe; a = probability of guessing 'target', P_r = probability of classifying a related probe as a target based on phantom recollection.

Replicating the base rate effect obtained in Experiment 1, parameter b was greater in the high b condition than in the low b condition, $\Delta G^2_{(df=1)} = 10.76$, $p < .01$. In contrast, and as predicted, parameter a was not affected by the base rate manipulation, $\Delta G^2_{(df=1)} = 0.45$, $p = .50$.

Gist-based false memory (parameter G_r) was again significantly greater in the Strong Gist as compared to the Weak Gist condition. Last but not least, verbatim memory for targets (parameter V_t) was not significantly increased in the Strong Gist condition.

Discussion

Results were clear-cut. Phantom recollection was again observed for critical lures from DRM lists from which eight items had been studied but not for critical lures from lists from which only a single item had been studied. In addition, the typical gist-based false memory effect on G_r was obtained, with stronger gist-based false memory for critical lures from eight-item DRM lists than for critical lures from one-item lists. Verbatim memory was not affected, supporting our interpretation that the unexpected V_t effect in Experiment 1 was an artifact of the presentation procedure. As predicted, the base rate manipulation affected parameter b but did not affect parameter a. This finding suggests that the effect on parameter a in Experiment 1 was indeed due to the manipulation of the proportion of related probes.

Parameter heterogeneity

In the present research, we analyzed data that were aggregated across participants using multinomial models. In doing so, it is assumed that parameters are homogeneous across participants. Violations of this assumption may lead to erroneous rejection of models, and significant results of parameter comparisons may be mere artifacts of data aggregation (e.g., Klauer, 2006). We tested the homogeneity assumption and, where it was violated, we fitted latent-class hierarchical multinomial models (Klauer, 2006) that provide an extension of multinomial models accommodating parameter heterogeneity. This hierarchical framework adopts a finite-mixture approach: it is assumed that participants belong to a specified number of different latent classes, and parameters are assumed to be homogeneous within latent classes but are allowed to vary between latent classes. Thus, between-participant heterogeneity in parameter values is accounted for by modeling two or more latent classes of participants that differ with regard to one or more parameters.

Klauer (2006) also introduced statistics to compare the empirical variance-covariance matrix to the variance–covariance matrix predicted by the model. Similar to methods used in structural equation models, the (asymptotically χ^2-distributed) S_1 statistic provides a quantitative description of the discrepancy between the two matrices (for detail see Klauer, 2006). A model is said to give an adequate account of the heterogeneity in the data when the discrepancy is negligible (i.e., when the S_1 statistic does not exceed the critical value). If the test is rendered significant for a model that assumes parameter homogeneity—

that is, a one-class model—then the homogeneity assumption does not hold. In this case, strictly speaking, traditional analyses using aggregated data should not be performed because of the risks of biased parameter estimates and inflated alpha levels. These risks do not apply to the latent-class framework, which can therefore be used to model heterogeneous data.

In order to test whether the central finding—the effect on P_r—might have been an artifact of data aggregation, we applied latent-class hierarchical multinomial models.[3] These models provide descriptions of the heterogeneous data by dividing the data into a given number of latent classes with different sets of parameter estimates. In Experiment 1, parameters were heterogeneous across participants, $S_{1(df=45)} = 122.31$, $p < .001$. A latent-class hierarchical multinomial model (Klauer, 2006) with three classes provided an acceptable fit for the heterogeneity present in the data, $S_{1(df=26)} = 35.43$, $p = .10$. Consistent with the findings reported above, parameter P_r was greater in the Strong Gist condition than in the Weak Gist condition, $\Delta L_{(df=3)} = 15.23$, $p = .001$; P_r was greater than zero in the Strong Gist condition, $\Delta L_{(df=3)} = 16.75$, $p < .001$, but not in the Weak Gist condition, $\Delta L_{(df=3)} = 1.33$, $p = .72$, indicating that the above findings were not due to artifacts of data aggregation.

In Experiment 2, parameters were again heterogeneous across participants, $S_{1(df=45)} = 149.10$, $p < .001$. A hierarchical multinomial model with four latent classes provided a good fit for the heterogeneity present in the data, $S_{1(df=18)} = 8.44$, $p = .97$. Corroborating the above findings, parameter P_r was greater in the Strong Gist condition than in the Weak Gist condition, $\Delta L_{(df=4)} = 23.34$, $p < .001$; P_r was again greater than zero in the Strong Gist condition, $\Delta L_{(df=4)} = 23.32$, $p < .001$, but not in the Weak Gist condition, $\Delta L_{(df=4)} = 3.11$, $p = .54$.

Summing up, we found that parameter heterogeneity across participants was substantial in both experiments. Latent-class hierarchical multinomial models were able to account for this heterogeneity. The analyses reported above were repeated using the latent-class models, and the main results were confirmed in these control analyses. We conclude that parameter heterogeneity across participants was present but did not substantially bias the main results obtained from the aggregated data in the present research.

General discussion

Results from two experiments support the validity of the phantom recollection parameter P_r. Phantom recollection was evident only when eight DRM list items were studied but not when a single DRM list item was studied. This finding is comparable to that obtained by Brainerd et al. (2001) who obtained strong evidence for phantom recollection for critical lures when participants studied 14 items from each DRM list. We can therefore conclude that the simplified CR model as depicted in Fig. 2 is a valid measurement model for the assessment of the phantom recollection phenomenon.[4]

Guessing processes

Batchelder and Batchelder (2008) suggested relevant metacognitive strategies for the guessing process modeled by parameter a. Applied to the simplified CR model, their framework predicts that estimates of parameter a be sensitive to the ratio of target probes to related distracters on the test list. This prediction was confirmed in Experiment 1 where the ratio of targets to related probes was manipulated. Experiment 2 provided evidence for the discriminant validity of parameter a by showing that it is not affected by the proportion of unrelated distracters on the test list.

The framework also predicts that parameter a be sensitive to the relative verbatim memory strength for targets as compared to related probes. Specifically, the tendency to guess "target" (i.e., parameter a) is predicted to be smaller in conditions with greater values of verbatim memory. This prediction was confirmed in Experiments 5 and 6 of Stahl and Klauer (2008) where estimates of parameter a correlated with the ratio of the verbatim memory parameters V_t and V_r. The prediction is also confirmed in the present data, where levels of verbatim memory V_t were greater in Experiment 1 than in Experiment 2. As predicted from the metacognitive account, guessing parameter a was, on average, smaller in Experiment 1 than in Experiment 2.

3 Because the latent-class framework does not yet provide a means to compare parameters across different groups (i.e., between-subjects effects), it was not possible to test for the effects of the base-rate manipulations within this framework. We therefore only report latent-class analyses for the effects on P_r that are central to the present paper. In order to obtain an identifiable model for a single group, we set recollection-rejection parameters to zero. We deemed this restriction unproblematic because the V_r parameters were not significantly greater than zero in the present data. To simplify the exposition, we collapsed across the base-rate manipulation and report only a single latent-class analysis for each Experiment. We also performed separate analyses for each condition that confirmed that the effect on P_r was not an artifact of aggregation. Furthermore, to investigate whether the effects on the guessing parameters a and b were mere artifacts of data aggregation, we computed separate latent-class analyses for each of the between-subjects conditions of each experiment and compared mean parameter estimates descriptively. These analyses corroborated the findings reported above. Details of these analyses can be obtained from the first author.

4 One reviewer suggested an alternative model in which the P_r parameter would appear on the first branch, leading to a 'target' response with probability P_r. With probability $1 - P_r$, the same tree as shown in Fig. 1 would then apply. This alternative model yielded an acceptable fit and its parameter estimates were comparable to those obtained with the model shown in Fig. 2. Importantly, hypotheses tests yielded the same pattern of results for the P_r-first and the V_r-first models. The model depicted in Fig. 2 was preferred here because of its similarity to the original model by Brainerd et al. (2001). Note that, in both models, significant results regarding parameter G_r depend on restrictions on V_r (i.e., equating V_r across Gist Activation conditions, or fixing V_r to zero; both assumptions are theoretically reasonable and empirically supported by model fit). Alternative models in which V_r was left free to vary and G_r was assumed to be unaffected by the gist manipulation also yielded acceptable fits; this was true for both the P_r-first and the V_r-first variants. However, the assumption that G_r is not affected by a strong manipulation of gist-based false memory does not make sense theoretically, and our previous research has shown that it is also not supported empirically.

The old correlational and the new experimental evidence, taken together, provide strong support for the validity of the simplified CR model's parameter a. It is a measure of the process of guessing whether an item judged old in meaning is a target or a related distracter, given that information in memory (verbatim memory or recollection rejection) is not sufficient to discriminate between these possibilities.

So far, we have assumed that the guessing process described by parameter a is the same whether gist has been retrieved (i.e., in the $[1 - V] * G$ branches) or not (i.e., in the $[1 - V] * [1 - G] * b$ branches). Note that this is a simplifying assumption that may not hold under all circumstances. There is some evidence suggesting that, in the closely related Source Monitoring paradigm, a related assumption is violated under certain conditions (e.g., Meiser, Sattler, & Von Hecker, 2007). In the present studies, as well as in our previous research in the simplified CR paradigm, this assumption has proven to be tenable (see also Stahl & Klauer, 2008).[5]

Predictions were also derived from the Bayesian framework for parameter b that is a measure of guessing whether an item is old (i.e., is either a target or a related probe) or new (i.e., an unrelated distracter): parameter b should be sensitive to the proportion of unrelated probes on the test list. This prediction was confirmed in both experiments.

Note that the Bayesian framework makes predictions not about the actual proportions of targets, related probes, and unrelated probes, but about the subjective proportions perceived by the participant; it is the subjective proportions that form the basis for an individual's guessing strategies. Thus, because the subjective proportions are not known, a certain amount of uncertainty is inherent in the predictions derived on the basis of actual proportions. Despite this uncertainty, the predictions derived from the actual proportions were confirmed, suggesting that the actual proportions were acceptable approximations of the subjective proportions, at least for the present studies.

Previous research on base-rate sensitivity of guessing processes in recognition memory has obtained mixed results (e.g., Estes & Maddox, 1995; Heit, Brockdorff, & Lamberts, 2003; Rhodes & Jacoby, 2007); the expected effects of base rate have sometimes failed to appear. In the present studies, the effects have been obtained reliably with as little as 20 participants and only 36 observations per participant. The surprising stability of the present base-rate effects might be for two reasons. First, the model-based analyses used here might have been especially sensitive. Specifically, the model's guessing parameters might be more sensitive for the effects of specific and targeted manipulations than the false-alarm rate, which may represent a mixture of effects of several (memory and) guessing processes.

Second, the robust findings in the present studies might have been caused not by the base rates themselves as they were picked up by participants on-line at test, but might have been based instead on the information provided to participants about these ratios in the instructions. In other words, it is unclear whether participants' knowledge about the proportion of the different probe types was affected at all by the actual proportions, or whether it stemmed mainly from the experimenter's instructions. An interaction between both factors is also conceivable, with the information provided in the instructions alerting participants to the possibility of different proportions of probe types on the test list. This notion is supported by a recent study suggesting that participants' awareness of the structure of the test list might be necessary condition to observe effects on response criterion (Rhodes & Jacoby, 2007). Whether the present effects were due to actual base rates or due to information provided about base rates to participants beforehand is not central to the present findings. The crucial finding is that the guessing processes were affected by the information about base rates in the predicted manner.

Levels of recollection-rejection in the original and the simplified CR paradigm

Previous research using the original three-groups CR procedure and model introduced by Brainerd et al. (1999, 2001) have observed nonzero values of the recollection-rejection parameter in experiments in which the DRM method was used (e.g., Brainerd et al., 2001; Exp. 1). This is in contrast to our own findings obtained with the simplified CR paradigm (Stahl & Klauer, 2008), in which we have consistently found estimates of recollection rejection to be (nonsignificantly different from) zero. Although not directly relevant to the present findings, this discrepancy warrants brief discussion.

Several differences exist between the two sets of studies, including, first, the exact materials, procedures, and participants; second, the different memory tests administered (i.e., the original three-group, between-participants procedure versus the present within-participants memory test); and, third, the different models used to analyze the data. Either one of these differences, or some combination, could be responsible for the discrepancy regarding the recollection rejection parameters. At this point, we can only speculate about its cause, because direct empirical comparisons of the present paradigm with the original CR paradigm (Brainerd, 1999), or the repeated-measures procedure recently introduced by Brainerd et al. (2008), are not available. Such a comparison study, in which the first two potential causes could be held constant, could help answer the questions raised above.

A possible reason for the discrepancy is the difference between models; it can be discussed here without the need for empirical data. In the original CR model, it is assumed that, if a verbatim trace cannot be retrieved, but gist memory is available, a related distracter is accepted as 'old' with probability 1. In contrast, the simplified model assumes that, in the same situation, a related probe is classified as a target with probability a. In other words, leaving aside

[5] To provide empirical support for this argument, we fitted a model with different a-parameters for the $(1 - V) * G$ and $(1 - V) * (1 - G)*b$ branches (a_G and a_b). V_r was fixed to zero, and phantom recollection P_r was fixed zero in the weak gist condition. In the present studies, the a_G and a_b parameters could be set equal without significant loss of fit, $\Delta G^2(2) = 0.74$, $p = .69$, for Experiment 1, and $\Delta G^2(2) = 2.88$, $p = .24$, for Experiment 2.

the obvious difference in memory test formats (i.e., old/new decisions in the original CR paradigm versus classifications in the simplified CR paradigm), the original CR model can be considered a special case of the present model with $a = 1$. Our previous studies using the simplified CR model have shown that estimates of parameter a were consistently and significantly smaller than 1 (Stahl & Klauer, 2008). These findings suggest that—at least for the classification test format used in our studies—the assumption of $a = 1$ cannot be upheld. This observation lends support for the simplified CR model.

The simplified CR model can be used to derive predictions for the old/new test used in the original CR paradigm. To see this, note that participants' 'old' response in the original CR paradigm depends on the instruction condition as follows: Under the T instruction, participants would respond 'old' only to targets, whereas under the R instruction, they would respond 'old' only to related probes (under the T+R instruction, they would respond 'old' to both targets and related probes). Using this mapping of classifications onto 'old' responses, predictions can be derived from the simplified CR model for the original CR paradigm. If the simplified CR model is the 'true' model not only for the simplified paradigm but also for the original CR paradigm, then 'true' values of parameter a smaller than one should lead to distorted estimates of the other parameters.

We conducted a small simulation study in which we generated data from the simplified CR model (which is the 'true' model in this simulation), and fitted these data using the original CR model. Typical parameter values were used, and the value of parameter a was varied between zero and one. Results showed that (1) the parameters of the original CR model were biased away from the 'true' underlying parameters generated by the simplified CR model for all values of $a \neq .5$, and (2), V_r was overestimated for values of $a < .5$. Given that estimates of parameter a were consistently below .5 in all our previous studies, and assuming that the simplified CR model is the true model, these results suggest that the nonzero values of V_r that were found in studies that used the original CR paradigm might reflect distortions due model misspecification. Note, however, that at this point these are speculations, not firm conclusions. Note further that the original CR model is at a structural disadvantage because it is not straightforward to use the original CR model to simulate data for the simplified CR procedure; this implies that the simplified CR model cannot be put to the same test. Additional research is required to investigate this issue.

Phantom recollection in other paradigms

Phantom recollection—false recollection of non-presented lures—occurs not only in the DRM paradigm. Using a procedure similar to the simplified CR procedure, Kensinger and colleagues (Kensinger, Garoff-Eaton, & Schacter, 2007) have investigated the effects of emotion on memory for gist and surface detail. In their Experiment 1, participants were presented with pictures of emotionally negative and neutral objects on neutral backgrounds, and were later tested for their memory for these objects and backgrounds. At test, the same, similar, or new objects and backgrounds were presented, and participants indicated whether a probe object or background was the 'same' as presented, was 'similar', or 'new'. On the basis of separate memory indices for recognition of gist and detail, the authors concluded that both types of memory were increased for negative over neutral objects, whereas both types of memory were reduced for backgrounds presented with negative objects as compared to backgrounds presented with neutral objects.

Due to the similarity in procedures, the present model can be applied to the data from Kensinger et al. (2007). In our application of the simplified CR model to the data, we found substantial evidence for phantom recollection.[6] That is, participants falsely responded 'same' to object and background probes that were not studied but were similar to studied objects and backgrounds, and they did so more often than would be expected on the basis of familiarity and guessing processes alone. This finding illustrates that phantom recollection can be observed in memory paradigms in which no special procedures were applied to create high levels of gist-based false memory. It also exemplifies the additional insight that can be obtained with model-based analyses.

The above reanalysis also illustrates how the simplified CR paradigm can be applied to materials other than DRM word lists, namely pictures of objects and backgrounds. In principle, the simplified CR paradigm can be used with any type of material for which the three probe types—targets, related distracters, and unrelated distracters—can be generated. This also applies to more complex materials such as entire sentences. In a recent paper in this journal, Singer and Remillard (2008) investigated veridical and false memory for sentence probes that were taken from, could be inferred from, or could not be inferred from previously read texts. In a comparison of different models, the authors show that the phantom recollection process is necessary to account for performance in immediate as well as in delayed testing. These findings further demonstrate that phantom recollection may occur in a variety of paradigms, using different materials, and they show that special techniques such as the converging-associates method implemented in DRM lists to elicit high levels of false memory are not necessary to obtain phantom recollection.

Acknowledgments

The authors thank Stephan Stegt for providing the materials used in both experiments.

[6] Given that the distribution of responses were very similar for neutral and negative new stimuli, we decided to set guessing parameters equal across neutral and negative stimuli in order to be able to evaluate model fit. Model fit for the object data was not acceptable for the model depicted in Fig. 1 that lacks a phantom recollection parameter, $G^2(2) = 15.82$, $p < .05$. Model fit was good when a phantom-recollection parameter was added to the model, $G^2(1) = 0.18$, $p > .05$. Phantom recollection was estimated at $P_r = .29$; this estimate was significantly greater than zero, $\Delta G^2(1) = 15.66$, $p < .05$.

The research reported in this paper was supported by grant Kl 614/31-1 from the Deutsche Forschungsgemeinschaft to the second author.

Appendix A

A.1. Extending the simplified CR Model: Phantom recollection

A.1.1. Data representation

In the simplified CR paradigm, the memory test list consists of target probes, related probes, and unrelated probes. The participant is required to classify each probe as either a target, related, or unrelated probe. This yields the following 3 × 3 frequency table F,

		Response		
		t	r	u
Probe	T	F_{11}	F_{12}	F_{13}
	R	F_{21}	F_{22}	F_{23}
	U	F_{31}	F_{32}	F_{33}

where F_{ij} is the frequency of event E_{ij}, that is, of a response of type j to a probe of type i.

A.1.2. Model equations

A multinomial model is created by expressing the probability p_{ij} of an event E_{ij} as a function of latent parameters that represent the psychological processes that are to be measured. In a multinomial processing tree model, the probability of an event is expressed as the sum of the probabilities of the branches that lead to that event. The following model equations can thus be easily derived from Fig. 2. For target probes, the empirical probabilities are given by the following three equations:

$$\begin{aligned} p_{11} &= V_t + (1-V_t)G_t a + (1-V_t)(1-G_t)ba \\ p_{12} &= (1-V_t)G_t(1-a) + (1-V_t)(1-G_t)b(1-a) \\ p_{13} &= (1-V_t)(1-G_t)(1-b) \end{aligned} \quad (1)$$

For related probes, the empirical probabilities are modelled as follows:

$$\begin{aligned} p_{21} &= (1-V_r)P_r + (1-V_r)(1-P_r)G_r a \\ &\quad + (1-V_r)(1-P_r)(1-G_r)ba \\ p_{22} &= V_r + (1-V_r)(1-P_r)G_r(1-a) \\ &\quad + (1-V_r)(1-P_r)(1-G_r)b(1-a) \\ p_{23} &= (1-V_r)(1-P_r)(1-G_r)(1-b) \end{aligned} \quad (2)$$

For unrelated probes, the empirical probabilities are modelled as follows:

$$\begin{aligned} p_{31} &= ba \\ p_{32} &= b(1-a) \\ p_{33} &= 1-b \end{aligned} \quad (3)$$

Together, Eqs. (1)–(3) specify how the simplified CR model (Stahl & Klauer, 2008) is extended by an additional parameter P_r for phantom recollection of related probes. Note further that this model differs from the model introduced by Brainerd et al. (2001) in that there is no parameter for erroneous recollection rejection, E_t. From the perspective of a model including a parameter E_t, the present model can be seen as a special case in which it is set to zero (see also Stahl & Klauer, 2008).

A.2. Identifiability

The above model is not globally identified because it has seven parameters for six free empirical probabilities. More specifically, the model is not identified because, in the tree for related probes (Eq. (2)), three memory parameters are used to model two free empirical probabilities (note that the guessing parameters are identified from the responses to unrelated probes, as shown in Eq. (3)). Given that there are only two free empirical probabilities, only two of the three memory parameters can be estimated from the data of a single condition. For P_r to be identified, a restriction therefore needs to be imposed upon at least one of the memory parameters for related probes, V_r or G_r. Setting V_r to zero is the most straightforward restriction here that can be maintained as psychologically plausible in certain situations (see main text).

From Eqs. (1)–(3) it follows that P_r is a function not only of the p_{ij} but also of V_r. Parameter P_r is given by Eq. (4). G_r is also a function of V_r as shown in Eq. (5) (equations for the other parameters are given in Stahl & Klauer, 2008).

$$P_r = \frac{p_{32}(1-V_r) - p_{32}p_{23} - (p_{22}-V_r)(1-p_{33})}{p_{32}(1-V_r)} \quad (4)$$

$$G_r = \frac{(p_{22}-V_r)(1-p_{33}) + p_{23}(p_{32}-p_{32}/p_{33})}{(p_{22}-V_r)(1-p_{33}) + p_{23}p_{32}} \quad (5)$$

From Eqs. (4) and (5) it can be seen that both P_r and G_r are identified (i.e., they can be expressed as a function of the empirical probabilities p_{ij}) when the value of V_r is not estimated from the data but instead fixed to zero or some other constant. In case $V_r = 0$, Eqs. (4) and (5) simplify to:

$$P_r = p_{21} - \frac{p_{31}}{p_{32}}p_{22}, \quad \text{and}$$

$$G_r = \frac{p_{22}(1-p_{33}) + p_{23}(p_{32}-p_{32}/p_{33})}{p_{22}(1-p_{33}) + p_{23}p_{32}}$$

This demonstrates identifiability of P_r and G_r for the case $V_r = 0$.

Other possible solutions involve setting V_r equal to constants other than zero. In principle, it would also be possible to obtain an identifiable model by imposing a restriction on parameter G_r. However, given that both P_r and G_r are measures of phenomena of gist-based false memory, such a restriction would be problematic in most cases. For example, setting G_r equal to zero will rarely be justified psychologically in a condition in which substantial P_r is expected.

Another possible way to obtain an identifiable model is to equate parameters across conditions. That is, V_r (or G_r) can be set equal to the respective parameter in another condition in which it is identified. For example, when phantom recollection is thought to occur in a condition A (with parameters P_r and V_r), an additional condition B can be included in the design (with parameters P_r' and

V'_r) in which phantom recollection is likely to be negligible, $P'_r = 0$, but recollection rejection is at levels comparable to those in condition A. In this situation, the restriction $V_r = V'_r$ would be justified. This restriction would yield an identifiable model because the parameter V'_r in condition B is identified (as shown in Stahl and Klauer, 2008; Appendix, Eq. 7). In Eqs. (4) and (5), V_r can then be replaced by a constant value of V'_r as computed from the empirical probabilities obtained in condition B (or, alternatively, by the right side from Eq. 7 in Stahl & Klauer, 2008). In the resulting equations, the parameters P_r and G_r are expressed only in terms of empirical probabilities; this demonstrates their identifiability.

In the present studies, an identifiable model was obtained by setting the memory parameters equal across the levels of the base-rate factor, which was implemented as a between-subjects manipulation. As a result of this restriction, the memory parameters in the tree for related probes are identified, as will be shown next. Assume that Eq. (2) describes the data from level 1 of the base-rate factor, and that Eq. (2a), with

$$\begin{aligned} p'_{21} &= (1-V'_r)P'_r + (1-V'_r)(1-P'_r)G'_r a' \\ &\quad + (1-V'_r)(1-P'_r)(1-G'_r)b'a' \\ p'_{22} &= V'_r + (1-V'_r)(1-P'_r)G'_r(1-a') \\ &\quad + (1-V'_r)(1-P'_r)(1-G'_r)b'(1-a') \\ p'_{23} &= (1-V'_r)(1-P'_r)(1-G'_r)(1-b') \end{aligned} \tag{2a}$$

describes the data from level 2 of that factor. Setting the memory parameters equal across both levels amounts to setting $V'_r = V_r$, $G'_r = G_r$, and $P'_r = P_r$. Solving for P_r yields

$$P_r = \frac{p'_{32}(1-V_r) - p'_{32}p'_{23} - (p'_{22}-V_r)(1-p'_{33})}{p'_{32}(1-V_r)} \tag{4a}$$

After equating the right sides of Eqs. (4) and (4a), we can solve for V_r:

$$V_r = \frac{p_{32}(p'_{32}p'_{23} - p'_{22} + p'_{33}p'_{22}) - p'_{32}(p_{32}p_{23} - p_{22} + p_{33}p_{22})}{p'_{32}(1-p_{33}) - p_{32}(1-p'_{33})} \tag{6}$$

This demonstrates identifiability for V_r. In Eq. (4), V_r can now be replaced by the right side of Eq. (6) to yield an expression of P_r in terms of empirical probabilities. The identifiability of G_r can be shown analogously by replacing V_r in Eq. (5) by the right side of Eq. (6).

References

Batchelder, W. H., & Batchelder, E. (2008). Meta-cognitive guessing strategies in source monitoring. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metacognition*. Hillsdale, NJ: Erlbaum.

Brainerd, C. J., Reyna, V. F., Bellinge, H., & Myers, J. (2008). Conjoint recognition and the word-frequency mirror effect, submitted for publication.

Brainerd, C. J., Forrest, T. J., Karibian, D., & Reyna, V. F. (2006). Development of the false-memory illusion. *Developmental Psychology, 42*, 962–979.

Brainerd, C. J., Payne, D. G., Wright, R., & Reyna, V. F. (2003). Phantom recall. *Journal of Memory and Language, 48*, 445–467.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*, 164–169.

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

Brainerd, C. J., Stein, L. M., & Reyna, V. F. (1998). On the development of conscious and unconscious memory. *Developmental Psychology, 34*, 342–357.

Brainerd, C. J., & Wright, R. (2005). Forward association, backward association, and the false-memory illusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 554–567.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1075–1095.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 120–128.

Greene, R. L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 687–695.

Healy, A. F., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition, 6*, 544–553.

Heit, E., Brockdorff, N., & Lamberts, K. (2003). Adaptive changes of response criterion in recognition memory. *Psychonomic Bulletin & Review, 10*, 718–723.

Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science, 9*, 61–65.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language, 56*, 575–591.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71*, 7–31.

Meiser, T., Sattler, C., & Von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *Quarterly Journal of Experimental Psychology, 60*, 1015–1040.

Odegard, T. N., & Lampinen, J. M. (2005). Recollection rejection: Gist cueing of verbatim memory. *Memory & Cognition, 33*, 1422–1430.

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 305–320.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Seamon, J. G., Luo, C. R., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language, 46*, 323–340.

Singer, M., & Remillard, G. (2008). Veridical and false memory for text: A multiprocess analysis. *Journal of Memory and Language, 59*, 18–35.

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for hierarchical multinomial processing tree models. *Behavior Research Methods, 39*, 267–273.

Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of verbatim and gist memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 34*, 570–586.

Stegt, S. J. (2006). *Die Rolle impliziter assoziativer Reaktionen bei der Entstehung von Pseudoerinnerungen im DRM-Paradigma [The role of implicit associative responses in the generation of pseudomemories in the DRM paradigm]*. Berlin: Logos.

Wright, D. B., & Loftus, E. F. (1998). How misinformation alters memories. *Journal of Experimental Child Psychology, 71*, 155–164.