

No source memory for unrecognized items when implicit feedback is avoided

Simone Malejka¹ · Arndt Bröder¹

Published online: 15 September 2015
© Psychonomic Society, Inc. 2015

Abstract In a recent empirical study, Starns, Hicks, Brown, and Martin (*Memory & Cognition*, 36, 1–8 2008) collected source judgments for old items that participants had claimed to be new and found residual source discriminability depending on the old-new response bias. The authors interpreted their finding as evidence in favor of the bivariate signal-detection model, but against the two-high-threshold model of item/source memory. According to the latter, NEW responses only follow from the state of old-new uncertainty for which no source discrimination is possible, and the probability of entering this state is independent of the old-new response bias. However, when missed old items were presented for source discrimination, the participants could infer that the items had been previously studied. To test whether this implicit feedback led to second retrieval attempts and thus to source memory for presumably unrecognized items, we replicated Starns et al.'s (*Memory & Cognition*, 36, 1–8 2008) finding and compared their procedure to a procedure without such feedback. Our results challenge the conclusion

to abandon discrete processing in source memory; source memory for unrecognized items is probably an artifact of the procedure, by which implicit feedback prompts participants to reconsider their recognition judgment when asked to rate the source of old items in the absence of item memory.

Keywords Source memory · Recognition · Signal detection theory · Threshold models · Bayes factor

Can you remember where or when you came to know a piece of information although you do not remember the information itself? Mandler's (1980) famous butcher-on-the-bus example would sound rather odd if you were able to specify that the person on the bus is the butcher from the supermarket, after you had failed to identify the person as being familiar to you in the first place. The ability to recognize previously encountered information is referred to as *item memory* or *recognition memory*, whereas the ability to specify the origin of that information is referred to as *source memory*. The term *source* includes not only the spatiotemporal context, in which the information was acquired, but also physical characteristics of the information itself (Johnson, Hashtroudi, & Lindsay, 1993), providing a body of source information that can be completely, partially, or not at all encoded (Dodson, Holland, & Shimamura, 1998). Due to the close relationship of item and source information, a critical question concerns whether source memory can be accessed separately from or only jointly with item memory. If source information can be retrieved in the absence of item recognition, it should be possible to successfully discriminate the source of an unrecognized item.

Research on the relationship between item and source memory per se is intriguing for a deeper understanding

This research was supported by a grant to the second author (BR 2130/7-1) from the German Research Foundation (DFG). The empirical study was conducted while the first author was a visiting scholar at the University College London enjoying the great hospitality of David Shanks and his lab. The authors thank David Kellen and Jeff Starns for valuable comments on an earlier version of this article as well as Daniel Heck and Michael Lee for their help with the Bayes factors.

✉ Simone Malejka
malejka@uni-mannheim.de

¹ Department of Psychology, University of Mannheim, Mannheim, Germany

of human memory, but this branch of research also puts measurement models of recognition memory to the test. These models are important tools for quantifying recognition performance that confounds the genuine ability to discriminate items (i.e., a person's accuracy) with decision behavior in case of uncertainty (i.e., a person's response bias). In the course of selecting an appropriate model, an ongoing debate in the recognition literature arose to answer the question of whether the processes of interest should be dissociated with models assuming continuous memory strength (signal-detection theory; Green & Swets, 1966), discrete memory states (threshold theory; Snodgrass & Corwin, 1988), or a combination of a signal-detection process and a threshold process (dual-process theory; Yonelinas, 1994). In case of item/source memory, popular instantiations of the competing theories make different assumptions about the relationship of item and source memory. In particular, the bivariate model of *signal-detection theory* (SDT) assumes that memory strengths for item and source information can vary independently to some extent (although they are correlated; DeCarlo, 2003), whereas the *two-high-threshold* (2HT) model declares item recognition to be a prerequisite for source discrimination (Bayen, Murnane, & Erdfelder, 1996). Contributing to the debate, Starns, Hicks, Brown, and Martin (2008) tested source memory for unrecognized items and showed that these items could be attributed to the correct source with above-chance probability. The authors consequently interpreted their finding as supporting the SDT model, and as being incompatible with the 2HT model.

Other works have also tested qualitative predictions of the competing measurement models by investigating whether certain types of information retrieval are necessary for source discrimination. In a recent study, Addante, Ranganath, and Yonelinas (2012) tested source memory in the absence of item recollection. They showed that not only high-confidence ratings (associated with item recollection) but also low-confidence ratings (associated with familiarity) could lead to correct source discrimination. This finding—in the context of correlations with respective event-related potentials—was taken as evidence in favor of dual-process models, in which recognition judgments stem from a threshold recollection process or from a strength-based familiarity process. The study, however, is not critical for threshold theory, because the 2HT model does not specify whether the old-detection state is reached via recollection or familiarity.

In a different study, Cook, Marsh, and Hicks (2006) used paired-associate learning to test whether recalling a target plays a role in retrieving contextual information. Source discriminability after unsuccessful cued recall was at or close to chance, but could be raised above chance when cue and contextual information were bound during encoding (e.g.,

through multiple study opportunities or by highlighting the importance of the cue-context association). This finding, of course, does not demonstrate source memory for unrecalled targets, as remembering the source of the cue would be sufficient for a correct source judgment. Furthermore, even without cue-context binding, a semantically related but non-studied cue was shown to elicit correct source discrimination after failed target recall, as long as the cue could reinstate the encoding context (Ball, DeWitt, Knight, & Hicks 2014). A similar argument applies to a study by Kurilla and Westerman (2010) who used word-fragment completion followed by an old/new judgment and a source judgment after an OLD response. When participants failed to complete the fragment but correctly recognized the corresponding word, they were also able to specify the source with above-chance probability. Here, source memory for the fragment (acting as the cue) suffices to solve the source-discrimination task and is certainly consistent with threshold models of recognition. Source memory without item recognition, however, was not examined as only recognized items were prompted for their source.

In sum, it seems that none of the findings outlined above compromises the 2HT model except for the finding by Starns et al. (2008). However, in the following, we outline the item/source-memory test used by Starns et al. (2008) and conclude that the particular procedure may have elicited the finding of source memory after apparently absent item memory by providing participants with implicit feedback. Here it is also explained in more detail why the SDT model appears to successfully predict the data of Starns et al. (2008), whereas the 2HT model fails to do so. We then propose an alternative procedure to examine source memory for unrecognized items. To test the models' predictions for the original and the new procedure, an empirical study is reported that successfully replicates Starns et al.'s (2008) finding—but only when feedback is provided. Finally, the results are discussed in light of the continuous-discrete modeling debate.

Source-memory tests for unrecognized items: A testbed for measurement models

The standard procedure to test item/source memory entails the assumption that there is no source memory for unrecognized items. After studying items under two conditions (e.g., male vs. female speaker), participants decide whether an item is old or new (item-memory test: target vs. lure) and judge the source of items that they have claimed to be old (source-memory test: source A vs. source B). Item and source judgments can be made successively with binary response options or simultaneously with ternary response options: OLD-SOURCE A, OLD-SOURCE B, and NEW. In either procedure, source memory for unrecognized items is

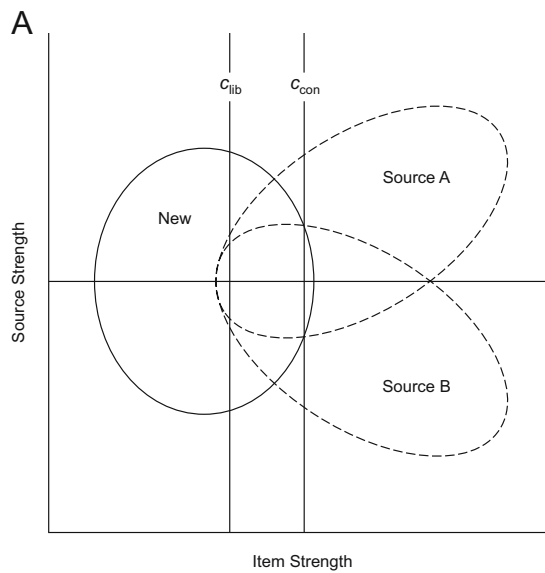


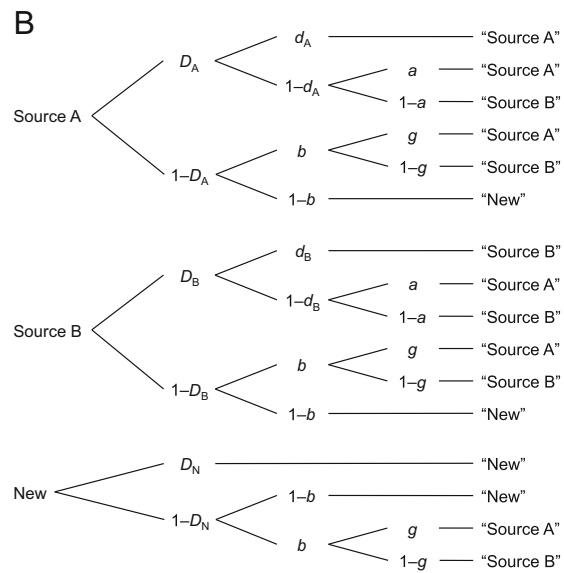
Fig. 1 (A) Bivariate signal-detection model of item/source memory. The ellipses represent 2-D slices through 3-D Gaussian distributions for new items, source-A items, and source-B items. The vertical lines indicate a liberal and a conservative response criterion on the item-strength dimension, whereas the horizontal line indicates an unbiased

not tested. In order to test it, Starns et al. (2008) conditionalized the source judgments not only on the recognition response, but also on the true old-new status of the item. In this conditional procedure, source judgments were required after OLD responses to any item type (i.e., hits and false alarms) and NEW responses to targets (i.e., misses) rather than after OLD responses only.¹

In three experiments implementing the conditional procedure, Starns et al. (2008) found above-chance source discriminability for unrecognized words when the item-memory test promoted conservative responding (by informing participants about a base rate of 25% targets). In Experiment 1 (small vs. large font) and Experiment 2 (pleasantness vs. imageability rating), the item-memory test preceded the source-memory test by asking for recognition judgments of all items and by asking for source judgments of hits, false alarms, and misses in two separate phases. The authors themselves raised two critical points concerning their experiments. First, items not recognized on the item-memory test may have been recognized on the delayed source-memory test due to repeated—and now successful—retrieval attempts.² Second, the base rate of targets was only

¹Note that DeCarlo (2003) refers to the standard procedure as the conditional procedure because source judgments are conditional on OLD responses.

²The same alternative explanation applies to a series of experiments conducted by Ceci, Fitneva, and Williams (2010), where context judgments of hits and misses led to apparent source memory for unrecognized pictures in children.



response criterion on the source-strength dimension. (B) Two-high-threshold model of item/source memory. Labels on the left refer to items presented on the item/source-memory test, whereas labels on the right refer to observed responses. Labels within the branches denote latent cognitive states that are entered with respective probabilities

instructed rather than actual, which may have promoted participants to artificially alter their responses in order to meet the response demand imposed by the instructed base rate. Therefore, in Experiment 3 (pleasantness vs. imageability rating), a source decision was required immediately after each recognition response to the same item and the base rate of targets was adopted as instructed. This matches the standard procedure of testing source memory as closely as possible and constitutes the most valid test of Starns et al.’s (2008) hypothesis—and thus the competing models outlined next—using the conditional procedure.

The bivariate SDT model of item/source memory assumes that each item presented at test produces evidence on an item-strength continuum and on a source-strength continuum (DeCarlo, 2003). An item’s evidence strength can be represented as a point in a two-dimensional coordinate system and varies according to three bivariate Gaussian distributions—one for each item type. The contours in Fig. 1A represent two-dimensional slices through the three-dimensional bivariate distributions with different means and variances in both dimensions. The distance between the distributions is a measure of the items’ discriminability, whereas the correlation between the distributions is influenced by the degree to which the applied task equally affects the different item types. By placing subjective response criteria along both dimensions, the evidence required to give an item response or a source response is determined. The distributional overlap marks the degree of uncertainty of how to respond.

For the conditional source-memory procedure, the response criterion c on the item-strength dimension of the SDT model is of primary interest (Starns et al., 2008). If this criterion is shifted to the left, it becomes liberal (c_{lib}) and only a small proportion of evidence values fall to its left side, where the source-A distribution and the source-B distribution almost completely overlap, resulting in poor source discriminability. In case of a shift to the right, the criterion becomes conservative (c_{con}) and a large proportion of the source-A distribution and the source-B distribution does not overlap, leading to above-chance source discriminability after incorrect NEW responses. Because items with higher evidence strength now result in NEW responses, source memory of unrecognized items should emerge.

In contrast, the 2HT model of item/source memory proposes that items can be in different discrete memory states (Bayen et al., 1996). Within each state, a transition to the next state follows with a certain probability. Figure 1B shows the 2HT model as a multinomial processing-tree model with three processing trees—one for each item type.³ For item recognition, the model proposes detection states for each item type reached with probability D and non-detection states reached with complementary probability $1-D$. If an item is detected as old or new, the respective response is given based on the memory for that event. If the item is not detected and thus the old-new uncertainty state is entered, an OLD response is guessed with probability b independent of any further memory information. In case of a hit and enough source evidence (d), a correct source response is given. In case of source uncertainty ($1-d$), SOURCE A is guessed with probability a . For false alarms, parameter g represents the probability of guessing SOURCE A. For misses or correct rejections, the standard source-memory procedure—for which the model was designed—does not ask for source judgments.

If you do ask for a source judgment after a miss in the conditional procedure, the 2HT model predicts that guessing the correct source—collapsed across sources to eliminate potential response tendencies toward one source—is at the 50%-chance level, because all information regarding the item such as its origin was lost when the old-new uncertainty state was entered. Therefore, according to the model, source discrimination for non-detected items is impossible or, to put it differently, source memory is limited to recognized items. After failed item detection ($1-D$), only item guessing with probability b and source guessing with probability

g is possible, because source discrimination is conditional on item detection (d only follows D). This is independent of the particular response bias b ($b_{con} < b_{lib}$).

Source memory for unrecognized items: A procedural artifact?

Irrespective of the specific details, all experiments reported by Starns et al. (2008) have one procedural detail in common. Whenever participants failed to recognize a target (i.e., they produced a miss during item recognition), the source of this item still had to be discriminated. The critical point here is that source judgments were required for hits, false alarms, and misses—but not for correct rejections. Hence, the procedure provides implicit feedback on the true old-new status of an item (see also Kurilla & Westerman, 2010, Footnote 2, and Klauer & Kellen, 2010). When only some of the NEW responses are followed by the source question, participants learn that these items must be old and that their recognition response was incorrect. Knowing that an unrecognized item was old may prompt participants to engage in second retrieval attempts. If these retrieval attempts are successful, they can also lead to correct source judgments.

Given the implicit-feedback interpretation, the finding by Starns et al. (2008) does not pose a threat to threshold theory and can be easily accounted for by the 2HT model. When second—and potentially more elaborate—retrieval from item memory is attempted, the state of old-new uncertainty may be left and seemingly lost information can be reactivated. If retrieval is successful, the old-detection state is entered with increased probability D and the item is not unrecognized anymore. Although the response in the item-recognition task cannot be changed retroactively, the source of the item can now be discriminated with probability d or not be discriminated with probability $1-d$.

To test whether implicit feedback actually leads to second retrieval attempts, we replicated Starns et al.'s (2008) Experiment 3, comparing their procedure to a procedure without such feedback. Our study followed the design as described by Starns et al. (2008). However, in addition to the conditional procedure, we implemented an unconditional procedure that asks for source judgments after every item independently of the recognition response and the true old-new status (i.e., after hits, false alarms, misses, and correct rejections on the item-recognition task). Although it may seem strange to participants to specify the source of lures, this is the only procedure that does not provide any feedback on the participants' performance during the test. Participants were instructed to make an educated guess of the most likely source if they had stated that the item is new. Later, misses could be analyzed separately. If there truly is source memory for unrecognized targets in the unconditional procedure, the additional source judgments of lures should not

³Multinomial processing-tree models are substantively motivated statistical models for categorical data that are used to measure latent cognitive processes (Batchelder & Riefer, 1999). The characteristic part of these models is the rooted tree structure with binary branching, after which cognitive processes follow each other directly and the success of the subsequent process depends on the success of the preceding process.

alter the finding of above-chance source memory, as long as participants are encouraged to attempt source retrieval after NEW responses. With regard to the second-retrieval interpretation, we predicted a replication of Starns et al.'s (2008) results—but only in the conditional procedure and not in the unconditional procedure.

Method

Design To investigate whether the finding of source memory for unrecognized items can be replicated in a procedure without implicit feedback, we implemented a 2×2 between-subjects design with source-memory procedure (conditional vs. unconditional) and old-new response bias (conservative vs. liberal) as independent variables.⁴ Materials, procedure, and analyses were matched as closely as possible with Starns et al.'s (2008) Experiment 3.

Materials A word pool of 200 English nouns was selected from the MRC psycholinguistics database (Coltheart, 1981). Each word was four to six letters in length and of moderate familiarity and concreteness, following the selection criteria specified by Starns et al. (2008). For each participant, the words were randomly assigned to serve as targets and lures. The study list consisted of 130 words, including 30 primacy and 30 recency buffer items. The length of the study list was chosen in order to decrease the performance on the item-recognition task, which should make the manipulation of the response bias more effective (cf. Starns et al., 2008). The test list consisted of 90 words. For a base rate of 25% targets (conservative response bias), 10 source-A items and 10 source-B items were interspersed with 70 lures, whereas for a base rate of 75% targets (liberal response bias), 35 source-A items and 35 source-B items were interspersed with 20 lures. Therefore, the actual base rates of targets were 22% and 78%.

Participants A total of 80 people associated with the University College London were recruited from the psychology department's participant pool and participated in return for a compensation of £4. The sample with 72 students (56 females) had a mean age of 24.13 years ($SD = 6.70$, range = 18–54).

Procedure Before the session started, participants were randomly assigned to the four experimental groups resulting

from a cross-classification of both between-subjects factors. Each group included 20 people, and each person was tested individually in a computer cubicle. The session consisted of a study phase, a brief distractor phase, and a test phase with the item judgments followed by the source judgments trial-by-trial. During study, participants were asked to rate 65 words according to Pleasantness (“How pleasant is the word?”) and 65 words according to Imageability (“How easy is it to imagine a referent to the word?”) on a scale from 1 (*very unpleasant/difficult*) to 5 (*very pleasant/easy*). The words appeared in random order. The 35 Pleasantness and 35 Imageability ratings in the middle of the list served as targets in the subsequent test phase. Prior to the first rating, participants were told that there were no right or wrong answers but that a memory test was going to follow. Each to-be-rated word remained on the computer screen until a rating was made, so that participants worked through the ratings at their own pace. After a 3-min picture task, participants were informed about the base rate of targets in the item-recognition task (25% vs. 75%) and the procedure of the source-discrimination task (conditional vs. unconditional). The conditional and the unconditional groups were told which respective items would be prompted for their source, and the participants were asked to give a source response after a NEW response that was most likely to be true for that item. At test, the list of 90 words was presented in random order. The participants indicated for every word whether it had appeared in the study phase and indicated for a subset of words—depending on their procedure—whether each had been rated for Pleasantness or Imageability. For the item and source decisions, response keys were labeled with stickers (“Y” for YES on the D-key, “N” for NO on the J-key, “P” for PLEASANTNESS on the C-key, and “I” for IMAGEABILITY on the N-key). At the end of the session, participants were thanked, debriefed, and compensated for participation.

Results

Recognition-memory performance To check the effectiveness of the response-bias manipulation, hit rates and false-alarm rates on the item-recognition task were analyzed (Table 1).⁵ In the conditional procedure, recognition hit rates for Pleasantness and Imageability items were higher in the 75%-bias conditions than in the 25%-bias conditions, replicating the performance reported by Starns et al. (2008). The same holds true in the unconditional procedure. Hit rates were submitted to a $2 \times 2 \times 2$ mixed analysis of

⁴The data collection started off as a 3×2 between-subjects design including the standard source-memory procedure, in which source memory is tested after every OLD response only. The factor level is not further mentioned because it is irrelevant for the present purpose (i.e., the event of discriminating the source of unrecognized items is not part of the standard procedure).

⁵Following Starns et al. (2008), all hit rates and false-alarm rates reported here were corrected according to Snodgrass and Corwin (1988) to eliminate hit rates of 1 and false-alarm rates of 0 for which the inverse of the cumulative standard Gaussian distribution is not defined.

Table 1 Mean (SE) performance on the item-recognition task and the source-discrimination task in the 25%-bias and 75%-bias conditions of Starns et al.'s Experiment 3 and of the conditional and unconditional procedures of the experiment reported here

Procedure	HR _P		HR _I		FAR		HR _S		FAR _S	
	25%	75%	25%	75%	25%	75%	25%	75%	25%	75%
Starns et al.	.76 (.04)	.89 (.02)	.75 (.03)	.88 (.02)	.05 (.01)	.11 (.03)	.72 (.03)	.74 (.02)	.27 (.03)	.23 (.03)
Conditional	.77 (.05)	.87 (.02)	.79 (.04)	.88 (.03)	.04 (.01)	.12 (.03)	.77 (.03)	.75 (.03)	.22 (.03)	.30 (.03)
Unconditional	.78 (.03)	.86 (.03)	.80 (.03)	.85 (.03)	.10 (.02)	.13 (.03)	.69 (.04)	.71 (.04)	.28 (.04)	.25 (.03)

Note. HR_P = hit rate on the item-recognition task for Pleasantness items only; HR_I = hit rate on the item-recognition task for Imageability items only; FAR = false-alarm rate on the item-recognition task; HR_S = hit rate on the source-discrimination task for Pleasantness items with a hit being a correct PLEASANTNESS response; FAR_S = false-alarm rate on the source-discrimination task for Imageability items with a false alarm being an incorrect PLEASANTNESS response. HRs and FARs are adjusted according to Snodgrass and Corwin (1988)

variance (ANOVA) with source as within-subject factor and response bias and source-memory procedure as between-subjects factors. The analysis revealed a main effect of response bias, $F(1, 76) = 6.98, p = .010, \eta^2 = .08$.⁶ The bias manipulation also led to higher false-alarm rates in the 75%-bias conditions than in the 25%-bias conditions of the conditional and the unconditional procedure. This was supported by a significant main effect of response bias in a 2×2 between-subjects ANOVA, $F(1, 76) = 5.40, p = .023, \eta^2 = .06$. Therefore, it can be concluded that the manipulation of the response bias was effective, leading to more OLD responses to targets and lures when liberal responding was promoted.

To test whether the response-bias manipulation influenced memory accuracy, performance measures adjusted for response bias were examined (see Table 2). Following Starns et al. (2008), a discrimination index was calculated as $d_R = \sigma \cdot [z(\text{HR}) - z(\text{FAR})]$, where σ is the standard deviation of the target distribution equal to 1.25 and z is the inverse of the cumulative standard Gaussian distribution.⁷ In the conditional procedure, d_R was highly similar in the 25%-bias and the 75%-bias condition when calculated separately for Pleasantness and Imageability items. The same result was found in the unconditional procedure. The $2 \times 2 \times 2$ mixed ANOVA revealed no main effects or interactions, all $F(1, 76) \leq 2.54, p \geq .115, \eta^2 \leq .03$. The data thus indicate no differences in discriminability for different sources and no differences in discriminability for different response biases.

⁶The conventional significance level of $\alpha = .05$ was used for all statistical analyses. All p -values refer to one-tailed tests.

⁷Although the measure d_a is recommended in case of unequal variances between lure and target distribution (Simpson & Fitter, 1973), Starns et al. (2008) adjusted d' for an assumed standard deviation of the target distribution of 1.25 (as suggested by Ratcliff, Sheu, & Gronlund, 1992). Because all three measures— d' , d_R , and d_a —are linear transformations of one another and ANOVAs will lead to the same results, only d_R is reported.

In addition to the discrimination indices, measures of the response bias were obtained to test whether the base rates of 25% and 75% targets promoted conservative and liberal responding, respectively. The distance of the response criterion from the mean of the lure distribution was calculated as $\lambda = z(1 - \text{FAR})$. As expected, the values of λ in the conditional and the unconditional procedure were always higher in the 25%-bias condition than in the 75%-bias condition, and the 2×2 between-subjects ANOVA revealed a significant main effect of response bias, $F(1, 76) = 6.80, p = .011, \eta^2 = .07$. Therefore, the response-bias measures and discriminability measures indicate that the response-bias manipulation successfully altered the response tendency but left the recognition accuracy unaffected.

Source-memory performance To check whether the response-bias manipulation in the item-recognition task influenced source memory, hit rates and false-alarm rates on the source-discrimination task were obtained and source memory was analyzed as source discriminability adjusted for response bias (Table 1). Source hit rates were calculated with hits being correct source responses to Pleasantness items, whereas source false-alarm rates were calculated with false alarms being incorrect source responses to Imageability items. Hit rates did not differ between the 25%-bias and the 75%-bias conditions in either the conditional or the unconditional procedure as supported by a non-significant main effect of response bias, $F(1, 76) = 0.02, p = .892, \eta^2 < .01$. The same applies to the false-alarm rates, $F(1, 76) = 0.45, p = .507, \eta^2 = .01$.

The distance between the means of the source-A and the source-B distribution was calculated as $d_S = \sigma \cdot [z(\text{HR}) - z(\text{FAR})]$, where σ is the standard deviation of the Pleasantness-item distribution equal to 1 (Table 2). The source-discriminability measures did not differ between the 75%-bias and the 25%-bias conditions either in the conditional or in the unconditional source-memory procedure. The main effect of bias condition was not significant,

Table 2 Mean (SE) signal-detection measures for item memory and source memory in the 25%-bias and 75%-bias conditions of Starns et al.'s Experiment 3 and of the conditional and unconditional procedures of the experiment reported here

Procedure	d_p		d_I		λ		d_S	
	25%	75%	25%	75%	25%	75%	25%	75%
Starns et al.	2.65 (0.16)	3.00 (0.17)	2.63 (0.16)	2.96 (0.16)	1.74 (0.07)	1.38 (0.12)	1.23 (0.14)	1.49 (0.15)
Conditional	3.50 (0.24)	3.18 (0.20)	3.54 (0.23)	3.26 (0.24)	1.89 (0.11)	1.35 (0.13)	1.66 (0.20)	1.31 (0.14)
Unconditional	2.84 (0.23)	3.17 (0.26)	2.97 (0.21)	3.14 (0.26)	1.41 (0.12)	1.32 (0.13)	1.26 (0.19)	1.37 (0.20)

Note. $d_p = d_R$ for Pleasantness items; $d_I = d_R$ for Imageability items; $d_R =$ signal-detection measure d' for an assumed standard deviation of the target distribution of 1.25; $\lambda =$ response-bias measure; $d_S =$ signal-detection measure d' for an assumed standard deviation of the Pleasantness-item distribution of 1

$F(1, 76) = 0.44, p = .510, \eta^2 = .01$. Hence, overall source discriminability in the source-discrimination tasks seems to be unaffected by the bias manipulation in the item-recognition task.

Proportion of correct source attributions Figure 2 shows the proportions of correct source attributions for all targets (Fig. 2A) and for unrecognized targets (Fig. 2B). The proportions for all targets were above chance in all combinations of response-bias manipulation and source-memory procedure, all $t(19) \geq 6.27, p < .001, d \geq 1.40$. In order to analyze source memory for unrecognized targets, proportions of targets attributed to the correct source conditional on NEW responses in the item-recognition task were determined. To test the pairwise differences between the bias conditions, simple-effects analyses were conducted. The comparisons revealed that the proportion in the 25%-bias condition was higher than in the 75%-bias condition only in the conditional procedure, $F(1, 76) = 7.58, p = .007, \eta^2 = .09$. The difference in the unconditional procedure failed to reach statistical significance, $F(1, 76) = 1.07, p = .303, \eta^2 = .01$. For the critical test of above-chance source memory without item recognition, the scores

were then analyzed using null-hypothesis significance testing and Bayesian model comparison of null and alternative hypothesis. The null hypothesis states that the mean proportion of correct source attributions for unrecognized targets is .50 ($\mathcal{H}_0: p_S = .50$), whereas the directed alternative hypothesis states that the proportion is larger than .50 ($\mathcal{H}_1: p_S > .50$).

Using one-sample t -tests, the proportion of correct source attributions for unrecognized targets was found to be larger than .50 only in the 25%-bias condition of the conditional source-memory procedure, $t(19) = 4.17, p < .001, d = 0.93$. All other scores did not reach the level of statistical significance, all $t(19) \leq 0.67, p \geq .257, d \leq 0.15$. This replicates Starns et al.'s (2008) crucial finding of above-chance source memory in the absence of item memory only in the 25%-bias condition of the conditional procedure, but not in the 25%-bias condition of the unconditional procedure.

There are two main objections to the t -test results. First, because the data entering the analyses are probabilities, the information about the total number of missed targets is lost. Second, the question of primary interest is not whether we fail to reject the null hypothesis in the 25%-bias condition

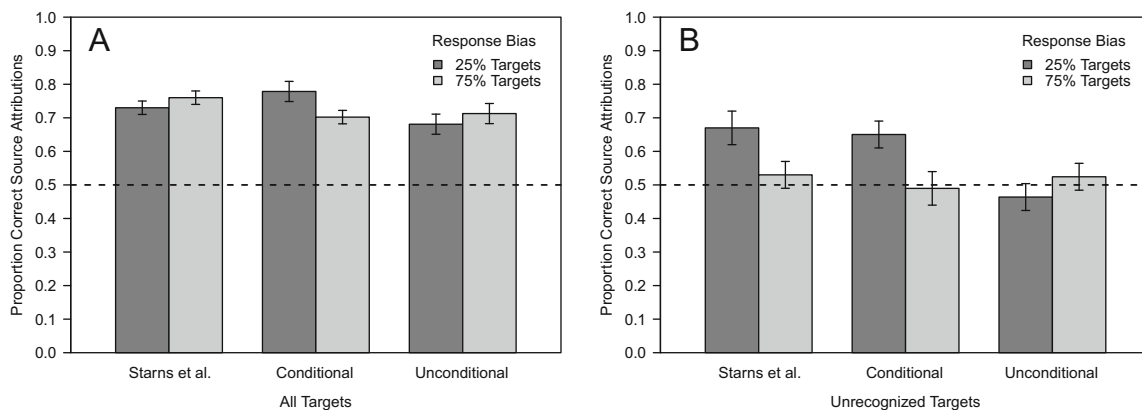


Fig. 2 Proportion and standard errors of correct source attributions for (A) all targets and (B) unrecognized targets in the 25%-bias and 75%-bias conditions as observed in Starns et al.'s (2008) Experiment 3

and in the conditional and unconditional procedures of the experiment reported here

of the unconditional procedure, but whether there is enough evidence to accept the null hypothesis. To account for both objections, we used Bayesian model comparison, which is a method of model selection based on the *Bayes factor* (BF; Jeffreys, 1961). The BF is calculated as the ratio of the marginal likelihood of the data under one hypothesis to the marginal likelihood of the data under another hypothesis, which denotes the relative extent to which the data support the first hypothesis over the second hypothesis (Kass & Raftery, 1995).

A latent-mixture model was used to compute the BFs (Lee & Wagenmakers, 2013). The model assumes that the response frequencies are generated by participants with latent membership in a source-guessing group or a source-knowledge group. The probability of making a successful source attribution is a rate, which is either a constant of .50 if the participants are in the guessing group or follows a beta distribution with $\alpha = \beta = 0.5$ truncated from below at .50 if the participants are in the knowledge group.⁸ To which group the participants belong is determined by a binary indicator variable (0 for the source-guessing group and 1 for the source-knowledge group). A Bernoulli prior with a probability of $\frac{1}{2}$ was used under the assumption that the group indicator is—a priori—equally likely to be 0 or 1. A BF can then be calculated directly as the posterior mean of the indicator variable divided by its complement.

The model to compute the BFs was applied to the data of each experimental condition separately. In the 25%-bias condition of the conditional procedure, the data were 110 times more likely to have occurred under the alternative hypothesis than under the null hypothesis, which can be considered as decisive evidence according to Jeffreys (1961). In the remaining three conditions, the data were 21, 12, and 6 times more likely to have occurred under the null hypothesis in the 75%-bias condition of the conditional procedure, the 25%-bias condition of the unconditional procedure, and the 75%-bias condition of the unconditional procedure, respectively. This denotes substantial to strong support for the null hypothesis.

Discussion

In the work reported here, source memory for apparently unrecognized items in the 25%-bias condition was replicated using Starns et al.'s (2008) procedure but not using a procedure without conditioning source judgments on participants' performance. The finding was supported by *t*-tests and BFs comparing the null hypothesis of chance performance to the alternative hypothesis of above-chance

performance. We therefore conclude that source memory for unrecognized items is probably an artifact of the conditional procedure that requires source judgments after misses but not after correct rejections. Source responses to hits, false alarms, and misses provide participants with selective feedback about their performance on targets and thus about the true old-new status of items. The feedback may prompt participants to rethink their incorrect item judgment when asked to discriminate the source of an apparently unrecognized item. This in turn may lead participants to a different memory state, in which item recognition was successful.

The opposing empirical predictions proposed by Starns et al. (2008) were thought to enable a model comparison without fitting competing models of item/source memory to data. The apparent finding of source memory for unrecognized items was interpreted as evidence against the 2HT model, because the model does not predict source memory for items in the old-new uncertainty state. The finding in light of the second-retrieval interpretation, however, does not constitute a threat to the 2HT model. A remaining weakness of the 2HT model is rather that the model does not naturally make different predictions for liberal and conservative response biases in the conditional procedure, whereas bivariate SDT does. More precisely, the effect of implicit feedback only manifested itself when the item-recognition task with 25% targets promoted conservative responding. Because there are usually more NEW responses in the 25%-bias condition than the 75%-bias condition, participants experience relatively fewer NEW responses followed by source questions as part of the total number of all NEW responses (calculated as the false-omission rate of the number of misses divided by all NEW responses). In the 25%-bias condition, the sudden source question following an incorrect NEW response after many correct NEW responses without the source question may make these trials more salient. Such rare but salient events may then encourage participants to spend additional time or effort to solve the task by retrieving the target. We admit that this explanation is post hoc, but it is open to empirical scrutiny. Moreover, when all current results are taken together, both rival models need to be altered post hoc; the 2HT model needs to explain the difference between bias conditions in the conditional procedure, whereas the SDT model needs to accommodate the results of the unconditional procedure.

If implicit feedback is responsible for the apparent finding of source memory for unrecognized items in the conditional procedure, an interesting follow-up question is how likely people are to succeed in recognizing an item that they initially failed to recognize, mainly because they were prompted to try again. On the one hand, when assuming stochastic independence of the success of two retrieval attempts, any probability of successful retrieval *D* (as stated in the 2HT model) will result in a higher overall retrieval

⁸The U-shaped beta distribution accounts for the log-linear correction by Snodgrass and Corwin (1988), where extreme proportions of correct source attributions are more difficult to obtain (Hautus, 1995). Using a uniform model prior leads to the same results.

probability if the attempt is repeated. The success rate for N attempts is $1-(1-D)^N$ which exceeds D whenever N and D are greater than 0. On the other hand, subsequent retrieval is more likely to be successful when more time is spent on the retrieval attempts. Research on reminiscence (i.e., recall of previously unrecalable items without relearning) has mostly focused on recall tests not recognition tests. However, increased net recall over successive recall efforts has also been shown for recognition memory (Bergstein & Erdelyi, 2008). This hypermnesic effect is generally not attributable to changes in memory that occur over time, but rather attributable to repeated testing (Roediger & Payne, 1982) or to additional retrieval time (Roediger & Thorpe, 1978). Furthermore, the finding of feedback-induced hypermnesia was linked to participants' performance expectancies (Klein, Loftus, & Fricker, 1994). Because feedback is always negative in the conditional procedure, rare but salient feedback in the 25%-bias condition may encourage participants to persist in their search for initially unrecognizable targets, whereas frequent feedback in the 75%-bias condition may rather discourage participants to spend more time or effort on second retrieval attempts.

Motivation also plays a part in the criticism of the unconditional procedure. One could question whether participants in the unconditional procedure exert less retrieval effort after a NEW response due to motivational factors or different strategies to solve the task. It cannot be ruled out that participants guess a source rather than indicate any tiny but crucial source detail that they might have been aware of. To prevent such behavior in the current study, participants were instructed to not simply guess but to always attempt retrieval of source information. Still, this potential motivational difference between the conditional and the unconditional procedure is interesting and could be easily tested by varying the proportion of NEW responses that are followed by the source question (e.g., equal proportions of misses and correct rejections prompting source judgments).

All procedural objections aside, the interpretation that implicit feedback created the results because of successful second retrieval attempts is not the only possible interpretation. Feedback may also give participants a reason to retrieve source details that they would not have retrieved without it. According to this interpretation, the implicit feedback informs participants that the item was studied, but the implicit feedback does not change that the item's evidence strength (as stated by the SDT model) was not strong enough to justify an OLD response. Participants may then infer that source information could be available, prompting them to attempt retrieval of source details and sometimes succeed at it. However, this view is somehow at odds with the conventional view of bivariate SDT, as the differential feedback is no longer defined as a confounding factor but redefined as a prerequisite of source memory after failed

item recognition. The source-detail interpretation entails the assumption that participants do *not* automatically consult the source-strength dimension if the item's evidence strength is below the response criterion on the item-strength dimension. However, participants would do so as soon as feedback is provided.

Although the work reported here will not resolve the modeling debate between discrete-state and continuous-strength models of recognition memory, we agree with Starns et al. (2008) that qualitative predictions and core assumptions of rival measurement models should be tested in extended paradigms in which no model fitting is necessary. Additional support for discrete-state models in an extended paradigm comes from a recent study by Harlow and Donaldson (2013), in which a novel source-memory task was introduced. This task measures source accuracy directly by calculating the degree to which a given source response deviates from the correct source response on a positional continuum. The authors found responses to be a mixture of accurate responses and guesses rather than responses of gradually decreasing accuracy. Hence, the nature of the processes in source memory may be discrete rather than continuous, and abandoning the 2HT model of item/source memory on the basis of apparent source memory for unrecognized items may be a premature conclusion.

Taken together, none of the studies outlined in the introduction directly examined source memory in the absence of item memory. Although the experiments conducted by Starns et al. (2008) appear to fill this gap, their interpretation is challenged by our study. Our findings rather support the idea that source information cannot be retrieved without item recognition. Hence, you may never be able to specify that the unfamiliar person on the bus is the butcher from the supermarket, unless you realize that you have made a mistake and that you do recognize the person after all.

References

- Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *Neuroimage*, *62*, 439–450.
- Ball, B. H., DeWitt, M. R., Knight, J. B., & Hicks, J. L. (2014). Encoding and retrieval processes involved in the access of source information in the absence of item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1271–1286.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree models. *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.
- Bergstein, J., & Erdelyi, M. (2008). Recognition hypermnesia: How to get it. *Memory*, *16*, 689–702.

- 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.
- Coltheart, M. (1981). The MRC psycholinguistics database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.
- Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 828–835.
- 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., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1121–1136.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, *27*, 46–51.
- Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review*, *20*, 318–325.
- Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Clarendon Press.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
- Klauer, K. C., & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review*, *17*, 465–478.
- Klein, S. B., Loftus, J., & Fricker, S. S. (1994). The effects of self-beliefs on repeated efforts to remember. *Social Cognition*, *12*, 249–261.
- Kurilla, B. P., & Westerman, D. L. (2010). Source memory for unidentified stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 398–410.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrences. *Psychological Review*, *87*, 252–271.
- Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). *Testing global memory models using ROC curves*. *Psychological Review*, *99*, 518–535.
- Roediger, H. L., III, & Payne, D. G. (1982). Hypernesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 66–72.
- Roediger, H. L., III, & Thorpe, L. A. (1978). The role of recall time in producing hypernesia. *Memory & Cognition*, *6*, 296–305.
- Simpson, A. J., & Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, *80*, 481–488.
- Snodgrass, J. G., & Corwin, J. L. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 30–50.
- 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.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341–1354.