Information accumulation on the item versus source test of source monitoring: Insights from diffusion modeling

Hilal Tanyas[1] · Julia V. Liss[1] · Beatrice G. Kuhlmann[1]

Abstract

Source monitoring involves attributing previous experiences (e.g., studied words as items) to their origins (e.g., screen positions as sources). The present study aimed toward a better understanding of temporal aspects of item and source processing. Participants made source decisions for recognized items either in succession (i.e., the standard format) or in separate test blocks providing independent measures of item and source decision speed. Comparable speeds of item and source decision across the test formats would suggest a full separation between item and source processing, whereas different speeds would imply their (partial) temporal overlap. To test these alternatives, we used the drift rate parameter of the diffusion model (Ratcliff, *Psychological Review*, *85*, 59–108, 1978). We examined whether the drift rates, together with the other parameters, assessed separately for the item and source decision varied as a function of the test format. Threshold separation and nondecision time differed between the test formats, but item and source decision speeds represented by drift rates did not change significantly. Thus, despite facilitation on the source decision when the item decision was immediately followed by a test for source memory than when item and source were tested in separate blocks, findings did not suggest that source information already begins accumulating in the item test in the standard format. We discuss the temporal sequence of item and source processing in light of different assumptions about the contribution of familiarity and recollection.

Keywords Source monitoring · Source memory · Item memory · Diffusion model · Temporal sequence

Introduction

Do we first remember a piece of information itself (i.e., item) and then its context features (i.e., source) – for example, where, when, or how we learned it – or can the retrieval of both overlap to some extent? Three decades ago, Johnson et al. (1993) introduced the source-monitoring framework and outlined the set of memory and judgment processes involved in attributions of mental experiences to their sources. Accordingly, both recognizing previous experiences (item memory, i.e., old or new?) and identifying their contextual details (source memory, i.e., source A or source B?) can be described within the source-monitoring framework with varying levels of differentiation. While item recognition succeeds even at lower differentiation levels, source attribution relies on more complete information. The objective of the current study was to test the temporal sequence of item and source processing more closely. Specifically, we were interested in whether source processing already starts in parallel to item processing or only starts sequentially to the successful item retrieval.

To illustrate that less differentiated information becomes available at earlier stages of processing, Johnson et al. (1994) investigated the time-course functions for item and source memory in an internal-external source-monitoring paradigm, also referred to as *reality monitoring* (Johnson & Raye, 1981), assessing memory for imagined versus perceived items. They employed the response-signal technique (Reed, 1973, 1976) and manipulated the amount of time allowed for retrieval systematically across varied response lags in a test where item and source judgments were collected simultaneously (response options: "imagined," "perceived," or "new"). This is an established experimental method that was also applied in earlier time-course studies investigating the temporal availability of context information (e.g., Dosher, 1984; Dosher & Rosedale, 1991; Hancock, 2002). Concerning its relevance to the focus of interest, the

✉ Hilal Tanyas
hilal.tanyas@uni-mannheim.de

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

response-signal technique was employed to compare the time course of item recognition with that of modality recognition (a source-memory task, see Hintzman & Caulton, 1997) or associative recognition (Gronlund & Ratcliff, 1989), differing from source memory with nuances in which each item is paired with another item (not a source) at study. Both studies revealed earlier availability of item information. Crucially, Johnson et al. (1994) further benefited from model-based approaches considering the multitude of processes involved in source-monitoring decisions. They assessed item and source memory separately and corrected for guessing based on multinomial-model parameters (Batchelder & Riefer, 1990). Consistent with the source-monitoring framework, the results suggested earlier accessibility of item information than source information. In subsequent years, more rigorous tests on Johnson et al.'s (1994) data (Kinjo, 1998; McElree et al., 1999) raised concerns regarding the conclusiveness of the original findings. Following that, Kinjo (1998, Experiment 1) conducted a stronger test in a modified procedure with more response-lag conditions and still observed that item memory was accessed before source memory. Later on, Spaniol and Bayen (2002) also used the combination of multinomial modeling (Bayen et al., 1996) and the response-signal technique. However, the authors compared the time-course functions of item memory and source *guessing bias* (but not source memory). They observed that item memory became available before source guessing.

Measurement models characterizing the underlying processes of item and source judgments with different assumptions (e.g., threshold models vs. signal-detection models; see Bayen et al., 1996, and DeCarlo, 2003, respectively) also fostered the time-course research indirectly by probing discussion on whether there is source memory for unrecognized items (e.g., Malejka & Bröder, 2016; Starns et al., 2008; see also Fox & Osth, 2022, for an overview). However, these models do not posit a temporal ordering between the memorial information. Specifically, even though the two-high threshold multinomial model of source monitoring (Bayen et al., 1996) postulates source discrimination as contingent on successful item recognition, the order of the item and source memory parameter in the multinomial model branches does not postulate a serial ordering. Rather, both could occur simultaneously in these branches (cf. Batchelder & Riefer, 1999). Indeed, Johnson et al. (1994) underlined this possibility of the parallel retrieval of item and source information under the serially represented structure of the multinomial model of source monitoring. Further, the source-monitoring framework predicts that differentiation of different memory characteristics can occur at different rates. However, it does not claim a full separation such that the completion of one processing is necessary for the onset of another processing. Instead, their time course can closely intertwine (see Fig. 1B in Johnson et al., 1993).

Interestingly, this possible alternative of parallel processing of item and source information is completely unaddressed by the published literature. Even though it can be concluded from direct investigations of response time lags that source retrieval completes later than item retrieval (cf. Johnson et al., 1994), it is not possible to ascertain from combining the response-signal technique and multinomial modeling whether source information is retrieved serial to completed item processing or already started in parallel to item processing (i.e., item processing starts before source processing, but their retrieval courses overlap to some extent). More broadly speaking, time-course curves of the response-signal technique display the change of accuracy as a function of response time characterized by separate parameters, thus allowing us to measure whether the onset of one processing occurs before the onset of another processing. Yet, under this particular circumstance, these time-course curves are described for parameter estimates of memory processes, and memory accuracy is measured with multinomial threshold models. Crucially, these specific analyses only tap into whether source processing accumulated sufficiently to cross the source-discrimination threshold but cannot indicate whether it was already started during earlier stages. Further, restricting the time available for responding may increase the risk of altering the cognitive processes and, in particular, their mental organizations. Although the response-lag technique is an insightful temporal study design, we still deem it important to further pursue this line of research with different methods. In contrast to the extensive investigation of item and source accuracy performance, for example with experimental dissociations (e.g., Lindsay & Johnson, 1991), research on the *temporal* aspects of item and source processing is relatively scarce, and to our knowledge, has only been conducted directly with the response-signal technique thus far. Therefore, to fully explore the breadth of the time-course question, we should expand our analysis to include spontaneous (i.e., not temporally restricted) source retrieval and, thereby, consider promising alternative methodologies.

Recently, Tanyas and Kuhlmann (2023) tried to address the question of the seriality versus partial overlap of item and source memory with the mouse-tracking method (cf. Kieslich et al., 2019), which allows us to measure these retrieval processes dynamically as well as to outline their temporal development. The item (old or new?) and source tests (source A or source B?) were presented either consecutively for each recognized item (i.e., directly following each "old" response) as in the standard research of source monitoring, or the source test of the recognized items was presented as a separate block after the full completion of the item test (for a similar blocked test procedure in source monitoring, see Osth, Fox, et al., 2018a). Contrary to the blocked format which provides relatively more independent measures of item and source (serving as the baseline), participants

made their source decisions more straightforwardly than their item decisions in the standard format as evidenced by the smoother (less curved) source trajectories than the item trajectories. There are two alternative interpretations of this pattern. First, during the item test of the standard format, source information might have been retrieved parallel to item information as preparation for the ensuing source test. Second, rather than parallel processing, being already in the item recognition state might have rendered source retrieval more accessible. Carefully note that in the second scenario, source retrieval can be assumed to have operated in sequence, rather than in parallel, to item retrieval and still can explain the observed difference in the source trajectory pattern. Thus, the observed movement trajectories during the source tests are not conclusively indicative of serial or parallel item versus source processing. It may seem that the temporal sequence of item and source processing is far from being resolved, but these results clearly underline the close links of item and source retrieval courses. Next steps will be to consider different techniques that are better suited to a closer look at this fine-tuning association.

To conclude, a more thorough examination is needed to capture important nuances which might underly response times (RTs) and mouse trajectories in source-monitoring processes. Notably, responses from such a higher-order cognitive task as source monitoring may reflect different processes of which only some are relevant to the item versus source attribution specifically. To disentangle these latent processes, the diffusion model is a promising candidate and may open up new avenues to the time-course question in source monitoring.

Diffusion modeling in episodic memory research

When considered from the traditional viewpoint of cognitive psychology, mean RT performance is considered the index of mental chronometry (cf. Balota & Yap, 2011). As a consequence of this approach, information from a number of experimental trials is condensed into a single mean, resulting in loss of information and a missing common metric that also accounts for accuracy. The diffusion model (Ratcliff, 1978) is highly recommended to overcome these problems because it includes full distributions of RTs of correct and incorrect responses (e.g., Vandekerckhove & Tuerlinckx, 2007; Voss et al., 2013; Wagenmakers, 2009). It assumes that during a binary choice task, information accumulates continuously until one of two thresholds (i.e., alternative decisional outcomes) is reached (see Fig. 1). This decision process is driven by systematic and random influences. Based on the RT and accuracy data from all test items, the model provides separate parameters for the speed of information accumulation (i.e., the drift rate, parameter v), the amount of information considered in decision making (i.e., threshold separation, parameter a), possible a priori decision
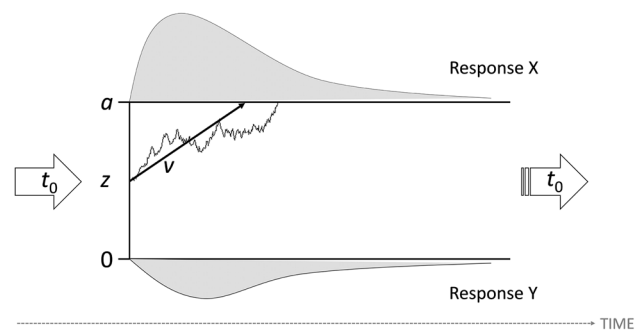

Fig. 1 Illustration of the decision process as proposed by the diffusion model (Ratcliff, 1978). Here, the upper and lower thresholds correspond to decisional outcomes (alternatively, correct and incorrect responses). The distance between the thresholds is represented by a. Information accumulation starts at z (here centered between the two thresholds) and continues over time with speed $|v|$ (denoted by the upward pointing arrow) until it reaches either of the two response alternatives. Random influences lead to unsteady fluctuations in the sample path. The duration of processes outside the decision process (e.g., encoding or response execution) are accounted for t_0. The response time distribution for choosing response X (response Y) is shown above (below) the respective threshold

biases (i.e., starting point, parameter z), and the duration of nondecisional processes (e.g., encoding and response execution, parameter t_0). In addition to these four key parameters, other parameters have also been added to the model over time such as to account for intertrial variability (i.e., parameters s_v, s_z, s_{t0}; see Ratcliff & Rouder, 1998; Ratcliff & Tuerlinckx, 2002) and differences in speed of response execution (parameter d; see Voss et al., 2010). Overall, the diffusion model allows researchers to understand whether – and especially in what ways – task performance can be explained by psychologically meaningful processes (Voss et al., 2013).

One important domain for the diffusion model is episodic memory tasks on which item recognition is assessed in binary response options. In his seminal study, Ratcliff (1978) introduced the diffusion model on the recognition memory paradigm. It then became a useful tool in recognition studies for several reasons such as to decompose age-related changes (e.g., McKoon & Ratcliff, 2012; Ratcliff et al., 2004, 2011; Ratcliff & McKoon, 2015; Spaniol et al., 2008), to investigate emotion-modulated memory (Bowen et al., 2016) and clinical disorders (White et al., 2010), and to enhance understanding of the strength-based mirror effect (Starns et al., 2012). The theoretical assumptions of the diffusion model are thus well met in recognition tests (Voss et al., 2013), and Arnold et al. (2015) further showed empirical validity of the diffusion model parameters for recognition memory. Most relevant to our research goal, Spaniol et al. (2006, Experiment 2) extended the use of the diffusion model to a two-choice source-monitoring task (also see Starns, 2014, for another application of the diffusion model to a source-memory task) to separately estimate the contributions of different processes

to episodic source retrieval in younger and older adults. They interpreted the drift rate in the source task as "the quality of the contextual information driving the decision process during retrieval" (p. 116). Importantly, the drift rate *expectedly* showed age differences in episodic but not semantic memory tasks, meaning that it was sensitive to the specific memory processing of interest. Inspired by this extension, here we employed the diffusion model to separate cognitive processes underlying both item and source decisions on the parameter level and, in particular, to better capture item versus source processing speed with the drift rate estimates.

The current experiment

In the standard sequential test of source monitoring (cf. Lindsay et al., 1991), item recognition for each trial is immediately followed by a source test. We aimed herein to investigate whether source decisions are reached after item decisions, compatible with this order of testing, or whether there can be some temporal overlap in item and source processing such that the latter is already started during the first item test step. Similar to the rationale of Tanyas and Kuhlmann (2023), we manipulated different test formats of source monitoring so that item and source information were either tested in immediate succession (i.e., the standard format) or temporally separated (blocked) for the recognized items (i.e., hits and false alarms). Thus, both test formats include an item test for all stimuli and a source test for items only judged as old. Critically, in the standard format, participants were informed in advance that they would be tested for the source immediately following each item recognition. In the blocked format, however, participants did not have prior knowledge about the upcoming source test block, and they were instructed to focus only on their item decisions during the item test block. Consequently, the blocked test format would provide a relatively more independent measure of item versus source decision speeds because the item and source tests are separated in time; thereby, they are more informative specifically about the duration of item versus source processing. Differences in decision speeds in the standard format with reference to the blocked format would then be informative regarding whether participants retrieved item and source in a sequenced or in a (partially) parallel way.

Of interest were RT data in relation to the accuracy of item and source test responses and parameter estimates for the diffusion model derived therefrom. We used the absolute values of parameter v as a measure of decision speed (of item vs. source processing, respectively) in each test. The faster the information accumulation, the higher the absolute drift rate. Carefully note that in comparisons across task conditions the drift rate maps onto task difficulty, such that easier tasks are associated with higher absolute drift rates (Lerche & Voss, 2019; Ratcliff & McKoon, 2008; Voss et al., 2004). Based

on higher differentiation and greater recollection demands in source memory (Johnson et al., 1993; Yonelinas, 1999), slower speed of information accumulation in the source test (parameter v_{source}) compared to the item test (parameter v_{item}) could be expected. Most importantly, however, we planned to compare the speed of one type of processing (i.e., item or source) with its pendant between the test formats. Our following hypotheses explain how these test format comparisons can inform us about the seriality or parallelism of item versus source processing:

H1 If we observe statistically comparable item and source decision speeds across the standard and the blocked test formats (i.e., no interaction of test format and memory type), this suggests a full separation (or temporal sequence) between item and source processing. Notably, in the blocked test format, while participants are responding old/new, they do not know yet whether (and when) there will be a test for source at all. Put differently, we do not give participants the chance to benefit from parallel retrieval of item and source in the item test block. Therefore, if item and source processing are sequential in the standard format, the speeds (i.e., of item and source) should always be the same as in the blocked format because this would always mean that the item is processed with its speed and subsequently the source is processed with its speed.

H2 By contrast, if item and source decision speeds differ by test format, our inference about temporal overlap would be based on the specific direction of condition differences. We would most plausibly expect the transfer of part (or all) of information accumulation from the source test of the standard format to its item test, which should be represented by slower item drift rates in the standard format compared to the blocked format. This would then suggest that source processing already started during the item test of the standard format. Consequently, we would also expect faster source processing in the standard format than the blocked format, indicating that part of the information accumulation in the source test must have been outsourced to the item test of the standard format.

As preregistered, we additionally explored whether the other parameters of the diffusion model also differ across the test formats in order to gain a better understanding of what composes the full RT distributions of item and source decisions in the standard sequential source-monitoring test.

Method

All materials and data together with our preregistration protocol are available online on the Open Science Framework. The preregistration protocol is available at https://osf.io/j9zwr/registrations. The experiment script and the results

(including the supplementary analyses) are openly available from https://osf.io/j9zwr/.

Participants

A power analysis for an *F* test conducted with the G*Power-3 software (Faul et al., 2007) indicated that a sample size of 30 per test format condition (*N* = 60) would provide a power of .80 to detect a medium-sized (*f* = .25) interaction between memory type and test format (α = .05, correlation among repeated measures = .10; see also Tanyas & Kuhlmann, 2023, for a similar logic). As explained in our hypotheses, the detection or rejection of this interaction is most relevant to deciding on the seriality versus parallelism of item and source processing.

Inclusion criteria to participate in the study were native fluency in English (learned before the age of 6 years); age (18–30 years); normal or corrected-to-normal vision; no diagnosis of mild cognitive impairment; no mental illness daily impact; no head injury that caused a knock-out for a period of time; no severe respiratory diseases (i.e., pneumonia or chronic obstructive pulmonary disease (COPD)); no medically diagnosed coronary artery or heart issues; no use of medication affecting cognition. We used Prolific's prescreening filters and additionally checked for these criteria before allowing participants to complete the full study. For the completed datasets, we also preregistered performance-based exclusion criteria that all participants should perform above-chance item memory (i.e., Hit rates > False alarm rates) and above-chance source memory (i.e., ACSIM score (average of the single-source conditional source identification measures, CSIM; cf. Murnane & Bayen, 1996) above .50). The reasoning behind that was that memory should drive most of the responses in the tests such that the drift rates tap into the speed of item versus source *memory* specifically. Thus, we recruited a total of 80 participants from the online recruitment platform Prolific (https://www.prolific.com/; also see Palan & Schitter, 2018) to meet the goal of analyzing data from a total of 60 participants. Nineteen participants were excluded from the data because they did meet the performance-based exclusion criteria. One participant took the study twice because of technical/internet problems and thus was also excluded. As reported later, we removed data from three participants based on our diffusion model analyses (see section Parameter estimation and model fit). Thus, the results reported are based on 57 participants (33 female, 23 male, one preferred not to indicate sex; M_{age}= 25.47 years, age range = 19–30 years). The experiment lasted approximately 30 min. Participants received payment according to the Prolific-set rate of £6/h.

Design

The design was a 2 (test format: the standard format vs. the blocked format) × 2 (memory type: item memory vs. source memory) mixed factorial design with memory type as a within-subjects and test format as a between-subjects factor. We also manipulated spatial position of study words (top vs. bottom of the screen) serving as the source manipulation as a within-subjects factor. However, we expected comparable item and source memory across sources and aggregated (as already planned in our preregistration) across spatial position for analyses (see Online Supplementary Material (OSM)).

Materials

We randomly selected 108 English nouns from the Toronto Word Pool (Friendly et al., 1982) after controlling for certain characteristics with the goal of selecting memorable items (imagery: ≥ 1.5 on a 7-point scale, concreteness: ≥ 2 on a 7-point scale, and Kucera-Francis frequency: ≥ 20). From this set, assignment of the words as study items (72 words) and distractors (36 words) as well as assignment of study items to the sources (50% on the top vs. the bottom of screen) were randomized anew across participants.

Procedure

We recruited participants on the platform Prolific, pre-filtering in accordance with our inclusion criteria. After seeing a detailed description and requirements of our study on Prolific, participants were redirected to OpenLab (https://open-lab.online/; Shevchenko, 2022) for the experimental task, which was programmed in an online study builder lab.js (based on HTML and JavaScript; see Henninger et al., 2022). The assignment to the test format conditions (the standard format vs. the blocked format) was randomized by OpenLab's urn function. After consenting, participants completed a demographic and health questionnaire, and we made sure that their responses were matched with Prolific's prescreening filters and thus checked our inclusion criteria again. If participants were deemed not eligible to participate, the session was terminated, and they received partial payment.

Eligible participants continued with the source-monitoring task. To increase memory-based responses in the later test, instructions emphasized before study that participants should learn both words (items) and their screen positions (sources) and that they would be informed later which exactly they will be tested on (cf. Tanyas & Kuhlmann, 2023). Before studying words, participants saw two fixed primacy buffer items (one on the top and one on the bottom with a randomized order for each participant) but

not as part of the words used in the source-monitoring test, and later they were presented in the practice test again along with two more distractor words. During study, 72 words were presented either on the top or on the bottom of the screen (random assignment of half of the items to each position) in a pseudorandom order with the restriction that there were no more than three consecutive repetitions in the same screen position. Each study item was shown in the respective position for 4 s, separated by a 500-ms inter-stimulus interval (a centered fixation cross and a blank screen, each lasting for 250 ms). Next, as a filler activity, participants verified simple math equations for 3 min. Finally, participants completed a self-paced source-monitoring test, designed according to their assigned test format condition. All stimuli were printed with 36-pt (corresponding to 48 px) Arial font in black against a white background throughout the experiment.

Participants in the standard format were informed that during their item decisions, if they indicated that a test trial was shown in the study phase before, their source memory for that trial would be tested immediately after (see Fig. 2A). In the blocked format, however, before the test session, participants were (truthfully) informed that only the words (not positions) matter for the responses here. We did this to maintain item test validity, as reasoned in the Introduction section. Thus, in this condition, participants were first questioned about whether the test trials were shown in the study phase or not, without being provided any information about the upcoming source test block yet. After the completion of the item test for all test trials, participants were then retested on the words they had judged to be "old" in the same order as on the item test and asked to indicate their studied positions (see Fig. 2B).

Participants in all conditions were instructed to respond as accurately and as fast as possible. At test, they were presented with a list that consisted of the items from both sources and new items (i.e., distractors), but this time all appeared centered one at a time. During the item test, the question "Have you seen this word before?" appeared in blue on the upper portion of the computer screen above the test
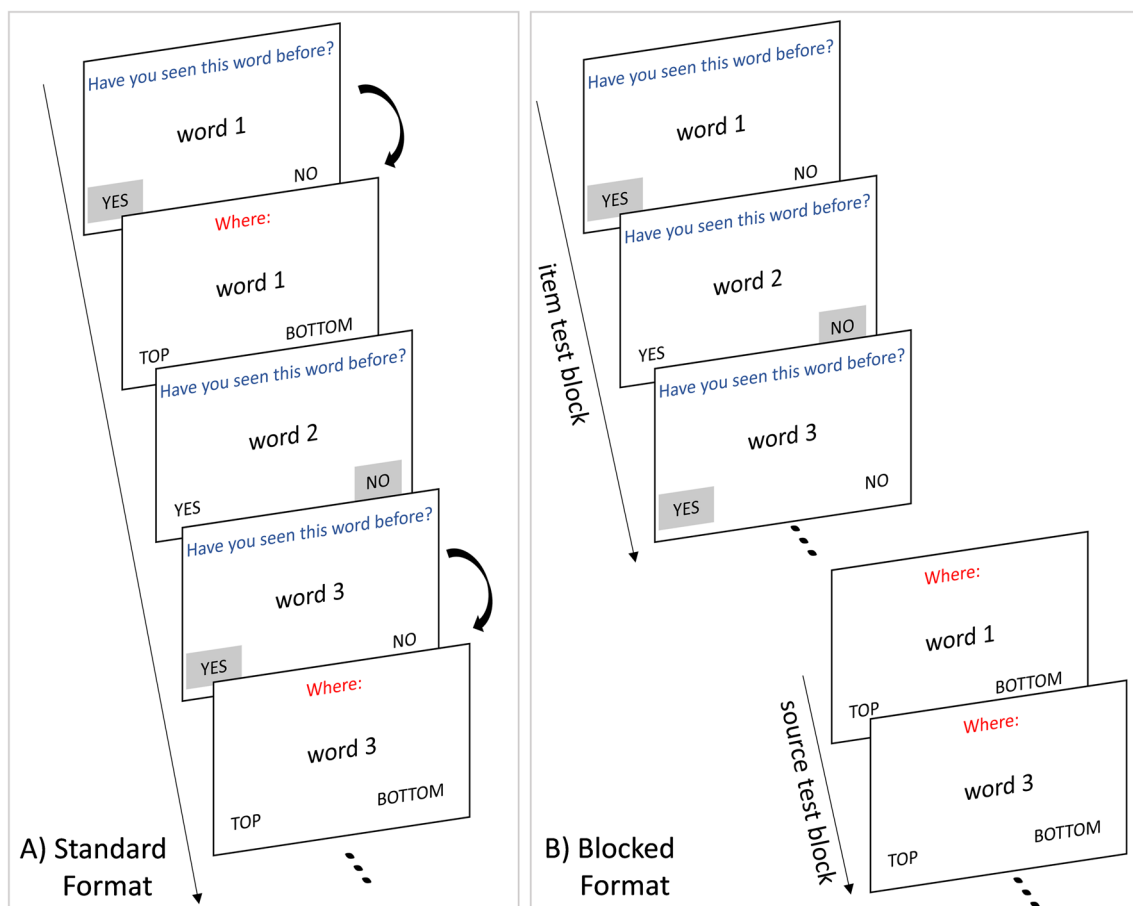

Fig. 2 Example visualizations of the test formats. (**A**) In the standard format, source decisions for each recognized test trial were collected in immediate succession to item decisions. (**B**) In the blocked format, source decisions for all recognized test trials were collected as a separate test block after the completion of the item tests

trials with the two response options. On the keyboard, "left arrow key" and "right arrow key" were assigned as "YES" and "NO," respectively. During the source test, however, the previous question was replaced by the word "Where:" and appeared in red, and both source options appeared side by side on the screen. In order to indicate source decisions, participants were required to press "up arrow key" or "down arrow key," representing "TOP" or "BOTTOM", respectively. These answer choices and key assignments were shown again on the test screen. We told them to simply guess if they could not remember whether and/or where the word was presented. Note that we needed targets to be able to assess source attributions, and more distractors are not particularly informative for our research question as for them a source cannot be retrieved. Therefore, we kept the same number of words for the categories "top," "bottom," and "new" (i.e., 36 for each). The presentation order of the test trials was also randomized by participants. Our lab.js scripts recorded response accuracy and RTs automatically.

Results

Parameter estimation and model fit

For the item test, we included both targets and lures in the analyses, and the thresholds of the diffusion model were linked to actual responses. Therefore, the upper and lower thresholds stand for "old" and "new" responses, respectively. However, for the source test, we restricted our analyses to only the items correctly identified as "old" because there cannot be source memory for lures as they were never presented with a source. The thresholds of the diffusion model were mapped to response accuracy. Thus, in the source test, the thresholds correspond to correct and incorrect source attributions given upon correct target detections. Considering the small trial number in our data, we used the maximum likelihood optimization criterion but with a strict outlier elimination procedure (following Lerche et al., 2017; Voss et al., 2013). Responses faster than 100 ms or slower than 4,000 ms were excluded from analyses (cf. Spaniol et al., 2006; Whelan, 2008). As an individualized elimination method, we additionally applied Tukey's outlier criterion (Tukey, 1977) separately for the item (considering the RT distribution based on both targets and lures) and source RTs to discard further possible contaminants. We removed trials that were more than three interquartile ranges below or more than three interquartile ranges above the third quartile of a participant's log-transformed RT distribution (e.g., Lerche, Neubauer, et al., 2018b). Prior to all analyses, we thus excluded a total of 2.21% of trials across participants.

Using the software fast-dm (Version 30.2; Voss & Voss, 2007; Voss et al., 2015), we fitted the diffusion model separately to each participant's data from the item test and the source test using the maximum likelihood optimization criterion. More specifically, for the item test, parameters comprised drift rates (v) for targets and lures, threshold separation (a), nondecision time (t_0), and relative starting point ($z_r = z/a$; $z_r > 0.5$ represents a bias towards the "old" response, whereas there is a bias towards the "new" response if $z_r < 0.5$). Given the sensitivity of the maximum likelihood optimization criterion towards fast outliers, we additionally estimated the intertrial variability of nondecision time (s_{t0}) to possibly avoid the negative effects of fast contaminants and to improve the estimation of the main parameters (Lerche & Voss, 2016). The intertrial variabilities of drift rate and starting point (s_v and s_{zr}), however, were fixed at zero for the sake of model parsimony and due to challenges associated with their estimation (Boehm et al., 2018; also see van Ravenzwaaij et al., 2017).

Note that the source test was conditional on the item test (i.e., an "old" response), indicating that the source trials entering the analyses were fewer than the item trials. We thus kept the model as simple as possible for the source test due to the restricted trial numbers (e.g., Lerche, Neubauer et al., 2018b; also see Lerche et al., 2017). More specifically, the upper and lower thresholds corresponded to correct and incorrect responses, respectively, and we assumed the relative starting point to be unbiased, fixing it at .5 (cf. Voss et al., 2013). Consistent with the item test, for the source test, we estimated s_{t0} once per participant and set s_v and s_{zr} to zero.

In total, we estimated ten parameters per participant (i.e., $v_{\text{itemtargets}}$, $v_{\text{itemlures}}$, v_{source}, a_{item}, a_{source}, t_{0item}, t_{0source}, $z_{r\text{item}}$, $s_{t0\text{item}}$, and $s_{t0\text{source}}$).[1] Two participants had to be excluded because parameter estimates could not be obtained due to very few source trials. One participant had a high proportion of too slow item responses rendering the parameter estimation difficult and was also removed from the results. We report the following analyses based on the remaining 57 participants.

An acceptable model fit is a prerequisite for analyzing and interpreting the diffusion model parameters. Figure 5 in the Appendix shows a graphical evaluation of model fit separately for the test formats by means of scatter plots. These scatter plots compare the accuracy rate (i.e., proportion of correct responses out of the total number of responses) and several RT quantiles of the behavioral data against the corresponding statistics predicted by the diffusion model based

---

[1] Interested readers can find the raw data as well as the dataset including individual estimates of the diffusion model parameters and the behavioral variables (i.e., accuracy rate and mean RTs) per participant in the OSF.

on the parameter estimates. The empirical and predicted values are described along the *x*- and *y*-axes, respectively. Each data point thus reflects one participant, and the discrete symbols accompanied by different colors refer to the item (along with item type) and source tests. Overall, data points are positioned tightly on or near the plots' main diagonal, indicating that the diffusion model provided a good account of the data of both groups.

Analyses of behavioral variables

Before further examining the diffusion model's parameter estimates, we first report empirical statistics for the behavioral variables' accuracy rate and mean RTs for the 57 participants included in the analyses. Mean accuracy rate and correct RTs are given in Fig. 3. Restricting to targets only, we performed separate 2 × 2 mixed ANOVAs using accuracy rate and mean RTs for correct responses with the within-subjects factor memory type and the between-subjects factor test format. The alpha level was set at .05, and we report partial eta squared (η_p^2) as the measure of effect size.

For accuracy rate, neither the main effects of test format nor memory type nor their interaction were significant (all *F*s < 1). Because accurate test responses may stem from memory or guessing, we also applied the multinomial processing tree (MPT) model of source monitoring (Bayen et al., 1996) to the present data as a more comprehensive analysis of the processes involved. Interested readers can find these more fine-grained accuracy analyses in the OSM.

In the analysis of correct mean RTs, the main effect of memory type, $F(1, 55) = 37.22$, $p < .001$, $\eta_p^2 = .40$, and the test format × memory type interaction, $F(1, 55) = 138.48$, $p < .001$, $\eta_p^2 = .72$, were significant, but not the main effect of test format, $F(1, 55) = 2.94$, $p = .092$, $\eta_p^2 = .05$. Relevant to our interest, the simple main effect analyses following up on the significant interaction revealed that correct mean RTs in the item test of the standard format were slower than in the item test of the blocked format, $F(1, 55) = 18.13$, $p < .001$, $\eta_p^2 = .25$. In contrast, correct mean RTs in the source test of the standard format were faster than in the source test of the blocked format, $F(1, 55) = 62.70$, $p < .001$, $\eta_p^2 = .53$.
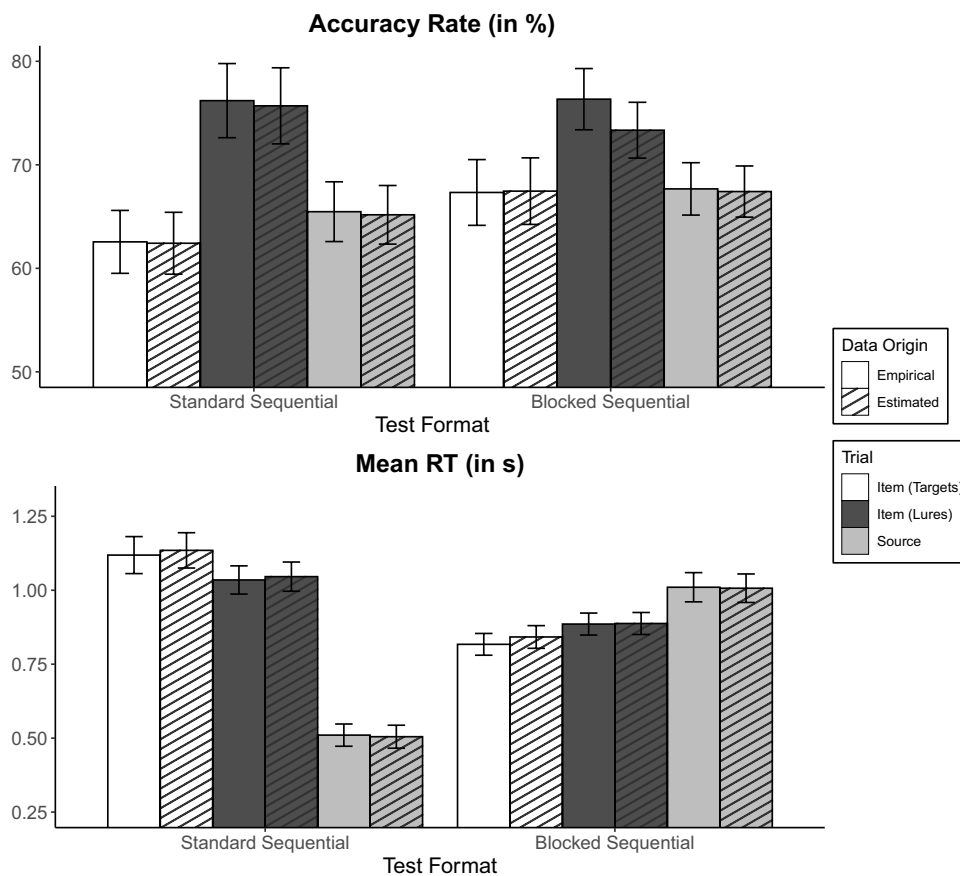

Fig. 3 Empirically observed mean accuracy rate and correct response times (RTs) across conditions with their estimated counterparts. Error bars represent standard errors. Mean RT = mean response time of correct responses

Table 1 Group mean estimates of the diffusion model parameters and the *t* tests for the comparison of test format

Parameter	Test format		*t* test		
	Standard format ($n = 27$)	Blocked format ($n = 30$)	$t(55)$	p	Cohen's d
$v_{\text{itemtargets}}$[a]	0.52 (0.08)	0.52 (0.11)	0.03	.975	0.01
$v_{\text{itemlures}}$[a]	0.94 (0.15)	1.27 (0.12)	1.73	.089	0.46
v_{source}[a]	0.66 (0.12)	0.53 (0.07)	−0.94	.349	−0.25
a_{item}	1.52 (0.08)	1.29 (0.06)	−2.42	.019	−0.64
a_{source}	1.16 (0.07)	1.47 (0.08)	2.87	.006	0.76
$t_{0\text{item}}$	0.58 (0.03)	0.51 (0.02)	−1.92	.060	−0.51
$t_{0\text{source}}$	0.18 (0.02)	0.50 (0.03)	8.36	< .001	2.22
$z_{r\text{item}}$	0.48 (0.02)	0.58 (0.02)	3.67	< .001	0.97
$s_{t0\text{item}}$	0.22 (0.03)	0.14 (0.03)	−1.98	.052	−0.53
$s_{t0\text{source}}$	0.07 (0.02)	0.16 (0.03)	2.24	.029	0.59

Note. Standard errors are presented in parentheses

[a]Prior to analyses, we calculated the absolute values of drift rate estimates to allow for the comparison of drift rates in terms of absolute size

Next, we describe our subsequent statistical analyses on individual estimates of the diffusion model parameters. Thereby, we can test whether the observed effects on item and source RTs reflect changes in the actual processing of item and source decision and/or in nondecisional aspects of the test responses (e.g., the motoric response).

Analyses of model parameters

Means of the estimates of the diffusion model parameters for conditions are presented in Table 1. Here, we examined three main diffusion model parameters (i.e., v, a, t_0) as the dependent variables and report inferential statistics on their estimated values. We conducted separate mixed ANOVAs using the individual parameter estimates of participants with the within-subjects factor memory type and the between-subjects factor test format. Bar plots for all three examined parameters are given in Fig. 4.

Drift rate (parameter v) indicates the direction and speed of information accumulation across all trials. Its sign is positive if the diffusion process reaches the upper threshold in the majority of trials and negative otherwise. Absolute drift rates, however, capture the speed of information accumulation independent of its correctness, with higher values representing faster accumulation (cf. Lerche & Voss, 2016). Our main interest was to understand whether participants already made part (or all) of the information accumulation for the source decision during the item test of the standard format (regardless of whether their decision processes mostly reached the correct or

incorrect threshold).[2] We thus submitted the absolute values of drift rates for the item (again, only targets) and source test to the 2×2 ANOVA. Neither the main effects of test format, $F < 1$, nor memory type, $F(1, 55) = 1.18$, $p = .282$, $\eta_p^2 = .02$, were significant. Importantly, test format did not interact with memory type, $F < 1$. This pattern is in line with our preregistered H1 supporting temporal sequence of item and source processing in the standard format. Item and source decision speeds did not differ by test format, such that the speed of information accumulation was statistically comparable when tested in succession or a blocked manner. Thus, there was no convincing evidence for a transfer of source processing to the item test in the standard format. In the following, we report additional exploratory analyses on the other diffusion model parameters.

Threshold separation (parameter a) is informative for how much information is required to decide on a response. For threshold separation, there was no significant main effect of test format, $F < 1$. However, the main effect of memory type, $F(1, 55) = 5.39$, $p = .024$, $\eta_p^2 = .09$, and the test format × memory type interaction were significant, $F(1, 55) = 51.32$, $p < .001$, $\eta_p^2 = .48$. Follow-up simple main effects analyses revealed that a larger amount of information was needed for an item response in the standard format compared to the blocked format, $F(1, 55) = 5.88$, $p = .019$, $\eta_p^2 = .10$. Notably, the respective difference was reversed in the source test such that less information was required in the standard format than in the blocked format, $F(1, 55) = 8.21$, $p = .006$, $\eta_p^2 = .13$. As threshold separation is part of the item and source decision making of interest to us, this pattern supports facilitation on the source decision by prior item retrieval (not significant for accumulation as captured by drift rates but significant for threshold).

Nondecision time (parameter t_0) estimates the remaining time outside the diffusion process such as encoding stimulus

---

[2] Note that we carefully considered performance-based accuracy as an initial step by calculating hits and false alarms on the item and source responses and excluding participants performing with poor accuracy (see Participants section). Thus, for the exclusions, accuracy in this study was assessed within the scope of source-monitoring perspective as summary measures on the participant-level. However, these were lenient criteria, and in our included dataset, there were still 14 participants whose item and/or source drift rates were negative, which indicates that their decision usually ended on the incorrect threshold. Put differently, their answers were not mostly driven by memory. We repeated our analyses without these participants for tapping into a closer inspection of memory-driven data. Of course, remaining participants' answers were not guessing-free either, but an exclusion of those 14 participants can be seen as a proxy to memory. Findings yielded similar patterns. Regarding drift rates, neither the main effects of test format nor memory type, Fs < 1, nor their interaction were significant, $F(1, 41) = 1.62$, $p = .210$, $\eta_p^2 = .04$.
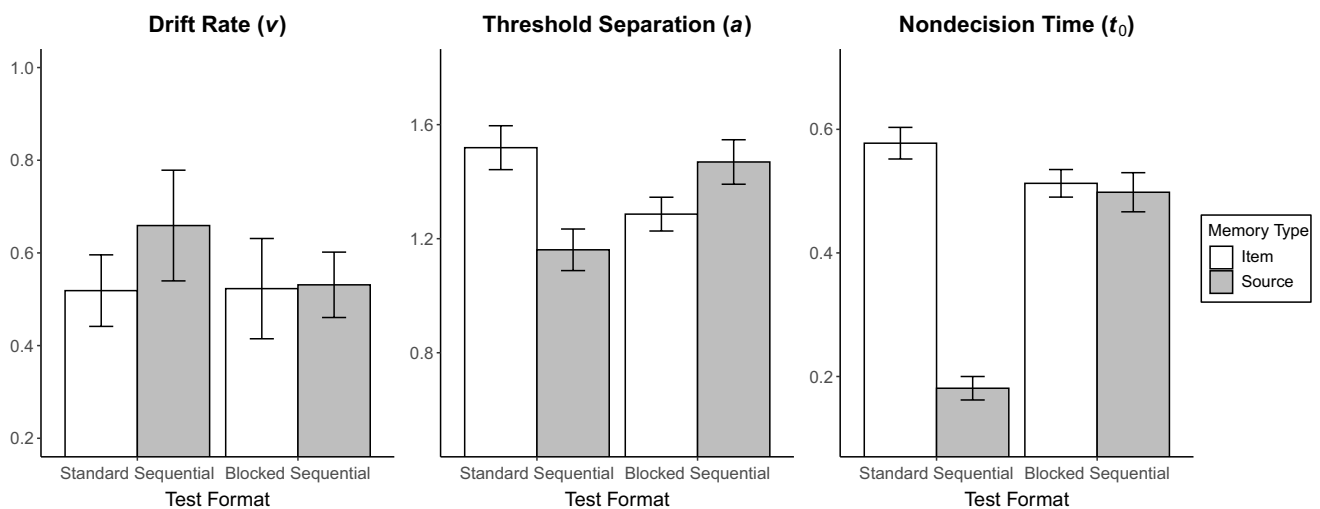
Fig. 4 Bar plots of the diffusion model parameters drift rate, threshold separation, and nondecision time. Error bars represent standard errors. For drift rates, absolute values are shown

and execution of the motor response. It is thus not informative about item and source processing per se but about differences in these additional demands between the two test formats. For nondecision time, the main effects of test format, $F(1, 55) = 16.02, p < .001, \eta_p^2 = .23$, and memory type, $F(1, 55) = 141.42, p < .001, \eta_p^2 = .72$, as well as their interaction, $F(1, 55) = 122.25, p < .001, \eta_p^2 = .69$, were significant. The simple main effects analyses further showed that during the item test, extradecisional processes took marginally longer in the standard format than in the blocked format, $F(1, 55) = 3.69, p = .060, \eta_p^2 = .06$. In contrast, during the source test, participants in the standard format had shorter nondecisional time compared to those in the blocked format, $F(1, 55) = 69.87, p < .001, \eta_p^2 = .56$. Overall, item responses were slowed down while source responses sped up in the standard format, and this seems not only due to response caution (threshold separation) but also due to nondecisional factors.

Discussion

The current research examined the temporal aspect of item and source processing in the standard sequential source-monitoring test by comparison with a blocked testing procedure. We collected source decisions for recognized trials either in immediate succession to item decisions as in the standard format or in a separate test block upon the completion of the item tests. The goal was to elucidate whether item and source processing are executed in sequence consistent with the order of standard testing (i.e., first item processing, then source processing) or whether there can be (partial)

temporal overlap between item and source processing during the item test of the standard format. To disentangle latent processes merged in raw RTs, while also considering accuracy, we applied the diffusion model (Ratcliff, 1978) analysis for each condition. Focusing primarily on the absolute drift rates, we compared the item and source decision speeds in the standard format with the blocked format to test the alternative time-courses. Although decision criteria and extradecisional processes showed differential effects across conditions, item and source decision speeds did not significantly differ by test format. On the behavioral level, participants knowing that they would be tested with an ensuing source test next upon their item decisions (i.e., an "old" response) gave slower item (but faster source) responses compared to those of the blocked format. However, these changes in RTs were unrelated to the speed of source information accumulation itself. Rather, our results point out response caution as a decision-level phenomenon and further underline the involvement of factors outside the decision process.

Given our preregistered hypotheses upon the interaction test of memory type and test format for the drift rates, we tentatively infer that there is no convincing evidence for temporal overlap between item and source. Accordingly, our results favored seriality, meaning that participants did not retrieve source information in parallel to item information when they were tested for their item memory. Otherwise, the transfer of information accumulation from the source test of the standard format to its item test should have led to a cost in information accumulation in the item test, and this should have been further supported by the reversed pattern in source decision speed.

The next question that could arise is why changes (or no considerable changes) especially in item decision speed

across different test formats indicate whether source processing is taking place in parallel or not. This question can be theoretically discussed benefiting from the underlying mechanisms that subserve item and source memory (i.e., familiarity- vs. recollection-driven processes). It is particularly noteworthy that in episodic memory research, there is an ongoing debate on the contributions of these operating mechanisms to item and source memory, and, importantly, these different assumptions also invite different interpretations of the current results. We thus additionally address an alternative explanation. On the one hand, arguing from the dual-process accounts of recognition (Yonelinas, 1999), one could claim that source memory primarily relies on non-automatic resource-consuming recollection, whereas item memory is merely familiarity-based, that is, driven by automatic processes of memory. Borrowing from this assumption, the reasoning behind our preregistered hypotheses can be, admittedly, argued against. By its common definition, automaticity refers to the aspects that are fast, governed by stimuli (not deliberate intentions), and less effortful (cf. Spaniol & Bayen, 2002). If we had observed significant changes in item decision speed in the standard format, it might have supported the notion that the serially following source response is interfering with prior item retrieval. However, the inference is less clear when we do not observe a significant change as in the current situation. One would intuitively assume that any parallel processing, if it occurs, is not shown by item decision speed because this might be an efficient processing without costs to the time spent on the item test. On the other hand, from the source-monitoring framework (Johnson et al., 1993), it is legitimate to assume that recollection also contributes to item memory. Empirical dissociations of item and source memory (e.g., Bayen et al., 1996; Lindsay & Johnson, 1991) do not necessarily imply that there is no resource dependency at all. Johnson (2005) emphasizes that other episodic memory tasks, such as item recognition, involve some processes that are not unique to source memory. In fact, item memory is not familiarity-based only, but rather, additionally requires the recollection of the experimental context (e.g., "Is this specific test item from the learning episode?"). More specifically, although recollection demands of item memory are not as specific as source memory, item recognition may not be effortless. Therefore, bearing on the role of slower recollection processes (e.g., Hintzman & Caulton, 1997; McElree et al., 1999), one could rightfully claim that if there was other processing taking place in parallel, it should have changed the item decision speed. Likewise, the start of – primarily recollection-based – source decisions in the item test should have been further represented by higher source drift rates in the standard format compared to the blocked format, but this prediction did not hold, either.

Thus, the current findings cannot present decisive evidence to distinguish between serial versus parallel processing of item and source because they cannot discard alternative time-course scenarios independent from the assumptions about the nature of the mechanisms underlying item and source decisions. However, the current results interestingly enough do show that in the often-employed standard source-monitoring task querying about the source following "old" responses on the item test does not necessarily change the speed of item information accumulation on this previous test. In other words, item information accumulation is robust to an ensuing source query. This means that item information accumulation can be well studied within a source-monitoring paradigm. Both of theoretical (regarding its automaticity status) and practical interest, it will be interesting to further test this robustness of item information accumulation under different conditions (e.g., different encoding conditions (cf. Lindsay & Johnson, 1991) or different test designs (cf. Fox & Osth, 2022)). Further, it seems important to study whether people can engage in source accumulation parallel to item accumulation and whether this may bring benefits to source memory. In continuing this research, we deem it important to conceive the time course of item and source processing may not be limited to the dyadic description (i.e., seriality vs. parallelism), but rather, the overlap possibly can vary along a continuum, which may even manifest differently under certain conditions or for different individuals (e.g., the differential age-related deficits on item and source memory (cf. Old & Naveh-Benjamin, 2008) may also alter their time-courses).

Drift rates are able to capture the changes on the memory tasks of interest (e.g., see Spaniol et al., 2006, for dissociating semantic and episodic memory drift rates), and thus can be quite informative for the temporal ordering of item and source processing with appropriate designs. Notably, McKoon and Ratcliff (2012) defined the drift rate in the memory domain as "the quality of the evidence from memory that drives the decision process" (p. 417). However, as also acknowledged by Spaniol et al. (2006), the diffusion model alone is incapable of addressing which component(s) of memory (i.e., encoding, maintenance (vs. forgetting), retrieval) "the quality" maps on. Despite its utility in summarizing latent processes (e.g., the rate of evidence accumulation), it is agnostic with regard to the explanations of these processes (e.g., underlying mechanisms that generate evidence accumulation in a memory task; for further discussion, see Osth, Jansson et al., 2018b). This important constraint can be enhanced by integrated models of memory and decision-making (e.g., adopting characteristics of memory representations together with the assumptions of accumulator models; see Cox & Shiffrin, 2017; Nosofsky et al., 2011; Osth et al., 2023). In particular, such a combined model framework has the potential to decompose the drift rate into properties that are relevant to theories of memory.

Another important finding worthy of further attention is that we observed substantial effects on threshold separation,

which gives further insight into the decision process underlying item and source responses. That is, the amount of information considered in decision making was significantly different between the test formats both for the item and source test. Compared to the blocked format, anticipating an immediate source test made the participants more conservative (i.e., setting higher thresholds) while making their item decisions. This difference was in the opposite direction for the source test, with smaller threshold separations in the standard format. This supports the notion that already being in the item recognition state requires less information to decide for a source response (cf. Tanyas & Kuhlmann, 2023). Technically speaking, prior item retrieval may just facilitate source attributions, thus leading to a closer distance between thresholds in the source test of the standard format, meaning that a smaller amount of evidence is necessary to accumulate in order to reach a threshold (i.e., make a source decision).

Overall, the distinct patterns in threshold separation suggests that our test format manipulation did really affect the decisional components underlying the item and source responses. Apart from threshold separation, further analyses of nondecision time showed that the test format manipulation not only yielded differences in decision settings but additionally affected nondecision-related factors. For example, comparisons with the blocked format indicated that in the standard format, nondecision time marginally increased during the item test but drastically decreased during the source test. A decrease in nondecision time is not surprising for the source test of the standard format since the same stimulus was tested again immediately after the "old" response. Thus, encoding of the stimulus in the source test should not have entailed as much time as in the blocked format, and this observed difference was most likely driven by perceptual encoding. At the same time, the intermixed presentation of the item and source test trials in the standard format means that participants also had to frequently switch between the keys assigned to the item and source responses. As a result, (marginally) increased nondecision time during the item test of the standard format can be mainly attributed to task preparation (cf. Schmitz & Voss, 2012) and motor activity – albeit being difficult to resolve precisely. It is thus clear that there are other factors underlying our test format manipulation that affect the speed of the responses without affecting the processing of the decision itself. The prominence of the drift rate enabled us to interpret the results that were corrected for nondecisional factors. Otherwise, it was

likely to observe the effects hidden in the mean (or median) RTs or confounded with related accuracy.

As a limitation of our study, we must acknowledge that the trial number in our dataset, albeit being typical for source-monitoring tasks, is conventionally considered small for the diffusion model analysis (but also see other instances, e.g., Lerche, Neubauer, et al., 2018b). Yet, previous simulation research investigating parameter recovery reveals that under certain conditions the diffusion model can offer reliable results even with small trial numbers (Lerche et al., 2017). Moreover, it is not always desirable to increase the number of trials because this may increase involvement of other processes (e.g., guessing) as memory would be overtaxed, and responses would no longer primarily result from the accumulation process which the diffusion model is assumed to measure (Lerche, Christmann, et al., 2018a). Here, we showed a successful application of the diffusion model to the source-monitoring paradigm by adopting the typical circumstances. In addition, our application shows that the diffusion model is applicable to a higher-order cognition which subsumes multiple processes, as is the case for the current source-monitoring study (see also Lerche, Christmann, et al., 2018a). We recommend episodic memory researchers to consider the feasibility and benefit of the diffusion-modeling approach, especially in RT studies. Finally, we acknowledge that our modeling approach implemented here is limited to separate treatment of the item and source RTs. While more sophisticated modeling techniques could simultaneously treat both test RTs and also model the cognitive processes involved in the source attribution (e.g., see supplemental MPT analyses in the OSM), we think that our experimental approach of manipulating and comparing multiple test condition RTs is insightful with the benefit of being a quite direct approach to studying this research question. Comparing response speed across experimental conditions to learn about seriality versus parallelism of processing has a long-standing tradition in cognitive psychology (Sternberg, 1969). Although, as discussed, one might argue whether strict seriality can indeed be inferred from our results, our drift rate analysis clearly shows that querying for the source right after the item response (as is standard in source monitoring research) does not change the core item processing. More sophisticated model-based analyses linking response speed to memory parameters would require making much more arguable assumptions about the nature of memory processes involved in source monitoring, which are heavily debated (cf. Fox & Osth, 2022).

Conclusion

Our study suggests that presenting item and source tests consecutively or in separate test blocks changed both decisional and nondecisional aspects of the item and source test responses. For the item response, we found a need for more information to decide and (marginally) increased nondecision time, resulting in a slower response in the standard format compared to the blocked format. However, when the source was queried immediately upon item detection, participants required less information for their source decisions and reduced nondecision time at this stage. Most importantly, although the way item and source memory are being tested affected other processes that confound overall RTs, the decision speeds matched with the drift rates did not change considerably across the test formats. A null effect in this interaction can be attributable to the lack of evidence for the temporal overlap of item and source processing, but it is also not warranted to assume a sharp separation between item and source retrieval even though they were probed in that order by the standard testing. Methodological progress can be made with more diagnostic models to rigorously assess direct connections between drift rates and memory characteristics (e.g., benefiting from process models). However, theoretical progress would likely come from the grounded assumptions of the source-monitoring framework, when the time-course question is reconciled more with the contributions of the different processes subserving item and source memory.

Appendix
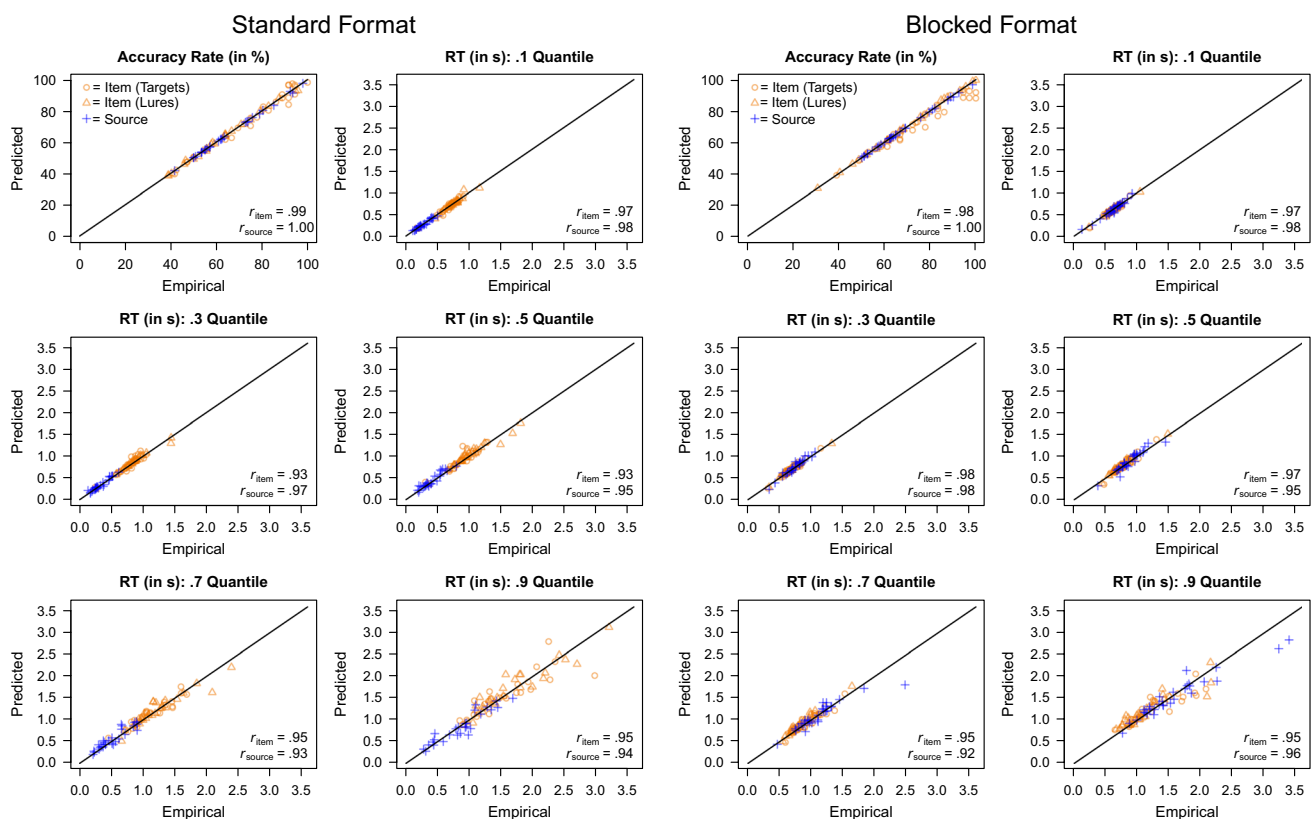
Fig. 5 Graphical displays of model fit. Concordance of the empirical and predicted statistics for the accuracy of responses and the .1, .3, .5, .7, and .9 quantile of correct response time distributions for each person in each condition. RT = response time; r_{item} = correlation between empirical and predicted statistics for item test; r_{source} = correlation between empirical and predicted statistics for source test

Declarations

References

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research, 79*, 882–898. https://doi.org/10.1007/s00426-014-0608-y

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20*(3), 160–166. https://doi.org/10.1177/09637214114088

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*(4), 548–564. https://doi.org/10.1037/0033-295X.97.4.548

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., Krypotos, A.-M., Lerche, V., Logan, G. D., Palmeri, T. J., van Ravenzwaaij, D., Servant, M., Singmann, H., Starns, J. J., Voss, A., Wiecki, T. V., Matzke, D., & Wagenmakers, E.-J. (2018). Estimating across-trial variability parameters of the diffusion decision model: Expert advice and recommendations. *Journal of Mathematical Psychology, 87*, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004

Bowen, H. J., Spaniol, J., Patel, R., & Voss, A. (2016). A diffusion model analysis of decision biases affecting delayed recognition of emotional stimuli. *PLOS ONE, 11*(1), 1–20. https://doi.org/10.1371/journal.pone.0146769

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review, 124*(6), 795–860. https://doi.org/10.1037/rev0000076

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology, 47*(3), 292–303. https://doi.org/10.1016/S0022-2496(03)00005-1

Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology, 16*(4), 519–555. https://doi.org/10.1016/0010-0285(84)90019-7

Dosher, B. A., & Rosedale, G. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. *Journal of Memory and Language, 30*(2), 125–160. https://doi.org/10.1016/0749-596X(91)90001-Z

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*(2), 175–191. https://doi.org/10.3758/BF03193146

Fox, J., & Osth, A. F. (2022). Does source memory exist for unrecognized items? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(2), 242–271. https://doi.org/10.1037/xlm0001111

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation, 14*(4), 375–399. https://doi.org/10.3758/BF03203275

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 846–858. https://doi.org/10.1037/0278-7393.15.5.846

Hancock, T. W. (2002). *How context changes the retrieval dynamics of a source* [Unpublished doctoral dissertation]. University of Georgia.

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods, 54*, 556–573. https://doi.org/10.3758/s13428-019-01283-5

Hintzman, D. L., & Caulton, D. A. (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language, 37*(1), 1–23. https://doi.org/10.1006/jmla.1997.2511

Johnson, M. K. (2005). The relation between source memory and episodic. *Psychology and Aging, 20*(3), 529–531. https://doi.org/10.1037/0882-7974.20.3.529

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67–85. https://doi.org/10.1037/0033-295X.88.1.67

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1409–1419. https://doi.org/10.1037/0278-7393.20.6.1409

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 111–130). Routledge.

Kinjo, H. (1998). *Recognition memory vs source memory: A comparison of their time-course in a speed–accuracy trade-off paradigm.* New York University.

Lerche, V., Christmann, U., & Voss, A. (2018a). Impact of context information on metaphor elaboration. *Experimental Psychology, 65*(6), 370–384. https://doi.org/10.1027/1618-3169/a000422

Lerche, V., Neubauer, A. B., & Voss, A. (2018b). Effects of implicit fear of failure on cognitive processing: A diffusion model analysis. *Motivation and Emotion, 42*, 386–402. https://doi.org/10.1007/s11031-018-9691-5

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology, 7*, 1324. https://doi.org/10.3389/fpsyg.2016.01324

Lerche, V., & Voss, A. (2019). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research, 83*, 1194–1209. https://doi.org/10.1007/s00426-017-0945-8

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods, 49*, 513–537. https://doi.org/10.3758/s13428-016-0740-2

Lindsay, D. S., & Johnson, M. K. (1991). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society, 29*(3), 203–205. https://doi.org/10.3758/BF03335235

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*(3), 297–318. https://doi.org/10.1016/0022-0965(91)90065-Z

Malejka, S., & Bröder, A. (2016). No source memory for unrecognized items when implicit feedback is avoided. *Memory & Cognition, 44*(1), 63–72. https://doi.org/10.3758/s13421-015-0549-8

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 563–582. https://doi.org/10.1037/0278-7393.25.3.563

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language, 66*(3), 416–437. https://doi.org/10.1016/j.jml.2011.12.001

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition, 24*(4), 417–428. https://doi.org/10.3758/BF03200931

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review, 118*(2), 280–315. https://doi.org/10.1037/a0022494

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018a). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language, 103*, 91–113. https://doi.org/10.1016/j.jml.2018.08.002

Osth, A. F., Jansson, A., Dennis, S., & Heathcote, A. (2018b). Modeling the dynamics of recognition memory testing with an integrated model of retrieval and decision-making. *Cognitive Psychology, 104*, 106–142. https://doi.org/10.1016/j.cogpsych.2018.04.002

Osth, A. F., Zhou, A., Lilburn, S. D., & Little, D. R. (2023). Novelty rejection in episodic memory. *Psychological Review, 130*(3), 720–769. https://doi.org/10.1037/rev0000407

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & McKoon, G. (2015). Aging effects in item and associative recognition memory for pictures and words. *Psychology and Aging, 30*(3), 669–674. https://doi.org/10.1037/pag0000030

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*(3), 438–481. https://doi.org/10.3758/BF03196302

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*(4), 408–424. https://doi.org/10.1016/j.jml.2003.11.002

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*(3), 464–487. https://doi.org/10.1037/a0023810

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science, 181*(4099), 574–576. https://doi.org/10.1126/science.181.4099.574

Reed, A. V. (1976). List length and the time course of recognition in immediate memory. *Memory & Cognition, 4*(1), 16–30. https://doi.org/10.3758/BF03213250

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 222–250. https://doi.org/10.1037/a0026003

Shevchenko, Y. (2022). Open Lab: A web application for running and sharing online experiments. *Behavior Research Methods, 54*, 3118–3125. https://doi.org/10.3758/s13428-021-01776-2

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 631–651. https://doi.org/10.1037/0278-7393.28.4.631

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 101–117. https://doi.org/10.1037/0278-7393.32.1.101

Spaniol, J., Voss, A., & Grady, C. L. (2008). Aging and emotional memory: Cognitive mechanisms underlying the positivity effect. *Psychology and Aging, 23*(4), 859–872. https://doi.org/10.1037/a0014218

Starns, J. J. (2014). Using response time modeling to distinguish memory and decision processes in recognition and source tasks. *Memory & Cognition, 42*, 1357–1372. https://doi.org/10.3758/s13421-014-0432-z

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), 1–8. https://doi.org/10.3758/MC.36.1.1

Starns, J. J., Ratcliff, R., & White, C. N. (2012). Diffusion model drift rates can be influenced by decision processes: An analysis of the strength-based mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1137–1151. https://doi.org/10.1037/a0028151

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Tanyas, H., & Kuhlmann, B. G. (2023). The temporal development of memory processes in source monitoring: An investigation with mouse tracking. *Psychonomic Bulletin & Review.* https://doi.org/10.3758/s13423-023-02289-z

Tukey, J. W. (1977). *Exploratory data analysis.* Reading: Addison-Wesley.

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review, 24*, 547–556. https://doi.org/10.3758/s13423-016-1081-y

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review, 14*(6), 1011–1026. https://doi.org/10.3758/BF03193087

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39*(4), 767–775. https://doi.org/10.3758/BF03192967

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*(7), 1206–1220. https://doi.org/10.3758/BF03196893

Voss, A., Voss, J., & Klauer, K. C. (2010). Separating response-execution bias from decision bias: Arguments for an additional parameter in Ratcliff's diffusion model. *British Journal of Mathematical and Statistical Psychology, 63*(3), 539–555. https://doi.org/10.1348/000711009X477581

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology, 6*, 336. https://doi.org/10.3389/fpsyg.2015.00336

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60*, 385–402. https://doi.org/10.1027/1618-3169/a000218

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*(5), 641–671. https://doi.org/10.1080/09541440802205067

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*, 475–482. https://doi.org/10.1007/BF03395630

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*(1), 39–52. https://doi.org/10.1016/j.jmp.2010.01.004

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1415–1434. https://doi.org/10.1037/0278-7393.25.6.1415