The Speed and the Temporal Aspects of Item and Source Processing in Source Monitoring

---

Hilal Tanyaş

Inaugural Dissertation

Submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences in the Graduate School of Economic and Social Sciences at the University of Mannheim

Thesis Defense:
September 19, 2023

Supervisors:
Prof. Dr. Beatrice G. Kuhlmann
Prof. Dr. Edgar Erdfelder

Dean of the School of Social Sciences:
Prof. Dr. Michael Diehl

Thesis Reviewers:
Prof. Dr. Arndt Bröder
Prof. Dr. Thorsten Meiser

Defense Committee:
Prof. Dr. Arndt Bröder
Prof. Dr. Thorsten Meiser
Prof. Dr. Beatrice G. Kuhlmann
Prof. Dr. Edgar Erdfelder

Table of Contents

Summary

The act of remembering does not solely include recognizing an encountered past experience (i.e., item) but also encompasses the critical ability to identify its episodic context (i.e., source). *Source monitoring* refers to this cognitive mechanism and subsumes memory (of item and source) and judgment processes by which the original source of a mental experience is determined (Johnson et al., 1993). In the literature, research on source monitoring aiming to understand dissociations between item and source processing has predominantly focused on accuracy performance. However, the speed of item and source processing and, in particular, their dissociations in time have received less attention. The overarching goal of the present dissertation is to examine the speed and the temporal aspects of item and source processing while concurrently benefiting from different methodological approaches.

In separate experiments, I investigated whether source information is retrieved after completed item processing (i.e., seriality) or whether both can partially overlap in time (i.e., parallelism). In Manuscript 1, using mouse tracking, I assessed how item decisions and source decisions for recognized trials (i.e., items judged to be old) develop qualitatively over time if they are collected in immediate succession (as in the standard test of source monitoring) versus in separate test blocks (i.e., a novel blocked test procedure created for comparison purposes). In Manuscript 2, as a more sensitive technique to distinguish between seriality versus parallelism, I applied the longstanding additive-factor method to the standard test of source monitoring. On the basis of the selective influence manipulations on item and source latencies, I examined whether item and source retrieval are executed in strict sequence. In Manuscript 3, I used the diffusion model to gain a deeper understanding of processing speed by disentangling decisional and nondecisional processes. Focusing on the drift rates, I compared the item and source decision speeds between the standard and blocked tests on the parameter level. Overall, relying on different measures, the present dissertation aimed to address the temporal sequence of item and source processing dynamically (mouse tracking), in real-time (the additive-factor method), and via a formal modeling approach (diffusion modeling).

Manuscripts

This dissertation was conducted in the Center for Doctoral Studies in Social Sciences (CDSS) of the Graduate School of Economic and Social Sciences (GESS) at the University of Mannheim and is based on one published article and two manuscripts under review. The presented research projects were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant KU3329/1-1, project number 392884168.

The main text provides a summary of each manuscript and reconciles findings across three projects along with their strengths and limitations. The original manuscripts appended to the main text present the experimental procedure and comprehensive analyses as well as the respective published literature of the particular methodology used in each project.

Manuscript 1

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

Manuscript 2

Tanyas, H., Kuhlmann, B. G., & Erdfelder, E. (2023). *Testing the serial processing model of item and source retrieval: Applying the additive-factor method to source monitoring* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Manuscript 3

Tanyas, H., Liss, J. V., & Kuhlmann, B. G. (2023). *Information accumulation on the item versus source test of source monitoring: Insights from diffusion modeling* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

1 Introduction

"But will all quantitative treatment of mental processes be out of the question then? By no means! An important factor seemed to be susceptible to measurement: I refer to the time required for simple mental processes."

F. C. Donders (1868/1969, pp. 413-414),
the father of mental chronometry

In 1868, Franciscus Cornelis Donders published a research paper titled "On the Speed of Mental Processes" (see Donders, 1969, for English translation), and this work became a pioneer in experimental cognitive psychology (see Jensen, 2006, for a review). He used *reaction times* (or *response times*, RTs; but see Luce, 1986, for the suggested distinction) as an observable behavioral measure to understand unobservable mental processes and developed the "method of subtraction". Accordingly, a psychological task could be manipulated to insert or exclude a particular stage of processing. RTs measured under different tasks with and without this subprocess can be informative for the mental process of interest. Figure 1 demonstrates a simple illustration of Donders' method. Consider two experimental tasks: A "simple" task where participants are expected to press a button whenever a stimulus (e.g., light) appears and a relatively more complex "choice" task where participants are expected to select a response from alternative response options (e.g., each stimulus color is associated with a unique response key). When the time taken for the simple RT task is subtracted from the choice RT task, it provides insights into the mental processes occupied by making the choice. In sum, the key idea is to add or delete cognitive stages activated via different tasks and to measure RTs under these conditions to infer latencies of mental processes.

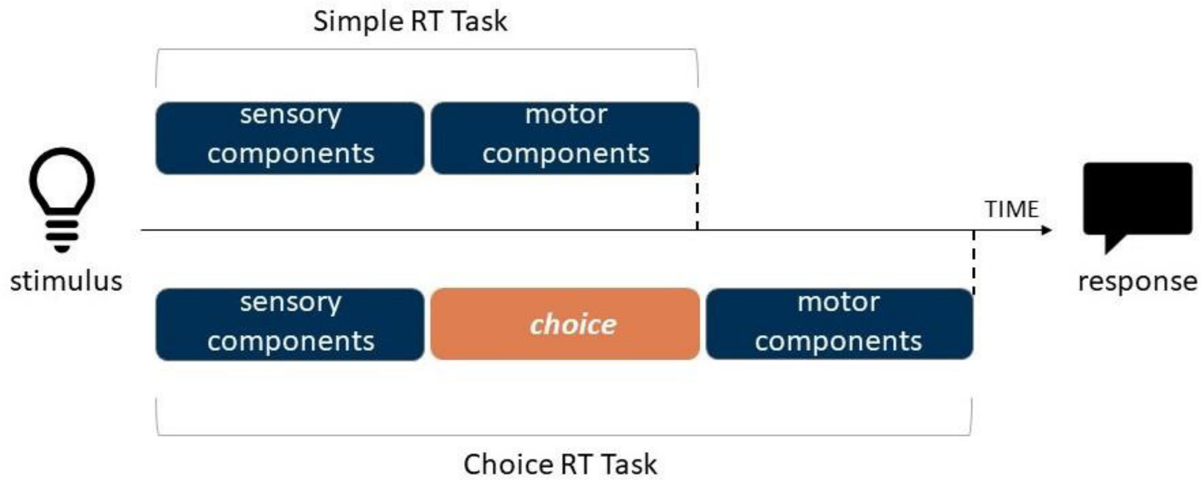


Figure 1. A simplified illustration of the method of subtraction. Note that the stages in this figure were prepared only for visual purposes. Their (relative) lengths herein do not represent their actual timing.

Critically, Donders' method is based on the strong assumptions that there is a processing chain from perceiving the stimulus to executing the response and that each component process starts after the completion of the preceding one (i.e., the assumption of seriality). Despite criticisms of the insertion procedure and its assumptions, this method is accepted as an innovation in the field (see Townsend, 1984) and influenced subsequent research (e.g., Sternberg, 1969). After the pioneering work of Donders, Wilhelm Wundt continued experimental research on RTs in the world's first psychology laboratory from the 1880s onwards (Jensen, 2006). Thus, the origins of RT studies indeed trace back to the early years of experimental psychology.

RTs—both in their simplest (raw RTs) or transformed form (e.g., log-RTs)—continue to be of relevance to this day and serve as a feasible and helpful index to understand human cognition in a variety of domains. By drawing on the advantages of mathematical modeling (e.g., Ratcliff, 1978), we can even further decompose them in a more sophisticated manner to utilize all information merged in RTs. As shown in Figure 1, there are several components underlying experimental tasks, and modeling based on RTs allows us to have better predictions on the specific components of decisions. However, with the advent of modern technology and the onset of the computer age, our scientific understanding of "time" is no longer limited to what RTs capture. We can additionally trace "temporal continuity" (Spivey & Dale, 2006, p. 207) during response selection, rather than merely collecting the outcome (i.e., the final decision and/or its speed). For this aim, one increasingly popular approach is mouse tracking. It is a process-tracing method and has already taken place in psychology to measure dynamics of processing (see Freeman et al. 2011; Schoemann et al., 2021, for reviews).

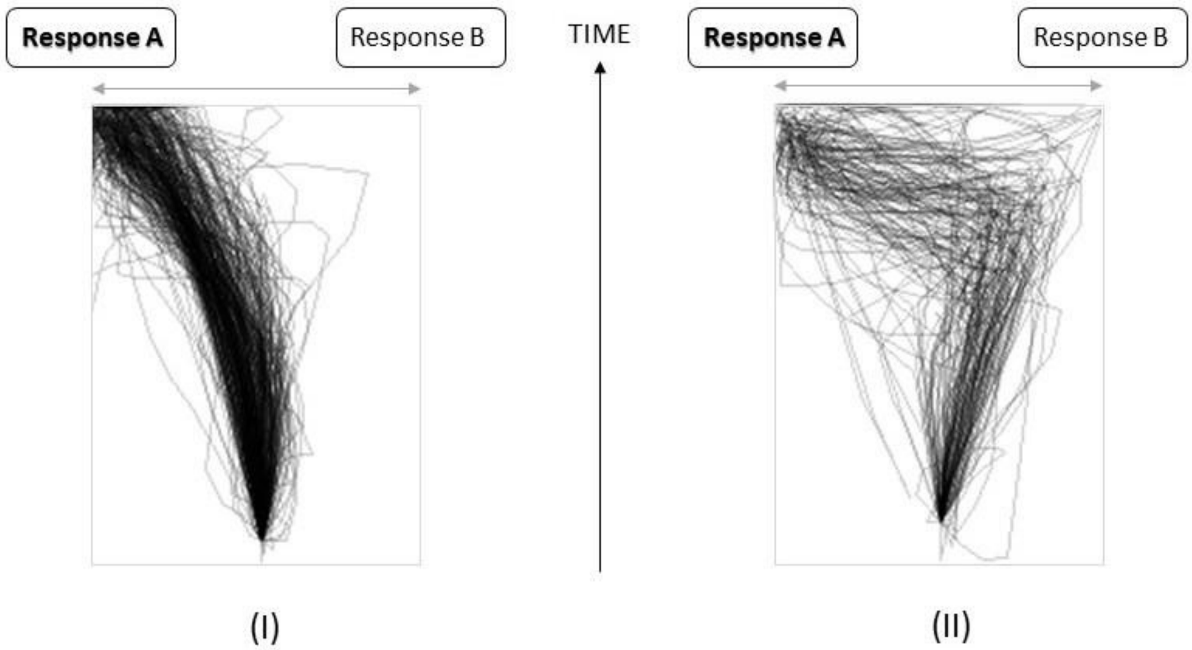


Figure 2. Example mouse trajectories in a binary-choice task. Note that these mouse trajectories were taken from the data collected for the first project of the current dissertation and categorized into two groups for ease of comparison (see D. Wulff et al., 2019).

To illustrate the role of mouse movements as a behavioral measure, I provide a concrete example in Figure 2. Consider a binary-choice task with the response options A and B separated on the screen, and suppose that panels I and II show the mouse trajectories of the correct responses of a participant (here: always option A). Even though all correct responses from both panels terminated on option A, the mouse trajectories that underlie these correct responses are qualitatively different, such that those in panel II are more curved compared to panel I. This indicates more deviation toward the nonchosen option (here: option B) for the trials in panel II. Importantly, while measures focusing on the outcome alone (e.g., RTs or response accuracy) cannot account for this activation between the response options, continuous measurement of mouse movements enables a closer look at how this outcome evolves over time (Freeman, 2018).

By presenting a holistic view, the current dissertation demonstrates how these measures, together with different methodological approaches, promote our knowledge about temporal aspects of cognitive processes. In the following sections, I introduce source monitoring as the central focus and outline the main motivation of the current dissertation, respectively.

2 The Source-Monitoring Framework

"A theoretical contribution of the reality/source monitoring approach is that it helped reconceptualize memory as involving an attribution about mental experiences rather than simply a revival of a stored representation."

M. K. Johnson (2005, p. 530),
a pioneer in source monitoring

Memories are derived from internal (e.g., obtained through imagination or thought processes) or external sources (e.g., being told by someone or seeing an object in a certain location). Based on certain attributes, we can infer the original source of an event or information (i.e., item). For instance, when you try to remember where you parked your car, this decision might be derived from remembering concrete cues, like park signboards and the color of the parking space, or from logical reasoning (e.g., I park my car in the same place every day). As such, to answer questions concerning our daily life, one must engage in source monitoring. Discriminating surrounding source features of central item information is of utmost importance for complete memory records and otherwise might pose severe consequences in real-life conditions (e.g., *where* you left your wallet, *who* advised you certain medical treatment).

In 1981, Johnson and Raye published a paper titled "Reality Monitoring", and they focused on how people discriminate information from internal and external sources. After almost a decade, Johnson et al. (1993) extended this work and presented the broadly defined term, *source monitoring*. In addition to distinguishing internal-external sources, source monitoring refers to all types of source discriminations, including internal-internal and external-external discriminations (also see Johnson & Mitchell, 2002). They defined *source* as “a variety of characteristics that, collectively, specify the conditions under which a memory is acquired (e.g., the spatial, temporal, and social context of the event; the media and modalities through which it was perceived)” (1993, p. 3). Furthermore, they formalized a *framework* to specify memory and judgment processes involved in source attributions. Since then, there is a growing body of episodic memory literature (see Lindsay, 2008, for a review) focusing on the source-monitoring framework (SMF).

Most relevant to my research goal, dissociations of item memory (a target event or information) and source memory (its episodic details) in source monitoring have been reported many times based on behavioral evidence (e.g., Bayen et al., 1996; Lindsay & Johnson, 1991). Given extensive research on the accuracy of responding, however, little is known about item and source retrieval *speed* and their temporal sequence. Before a thorough review of our current knowledge, I focus on how one tests source monitoring empirically.

2.1 The Source-Monitoring Paradigm

The source-monitoring paradigm consists of two phases: A study phase and a test phase. In the study phase, a number of items (e.g., words, sentences) are presented by at least two different sources (e.g., screen positions, agents). In the following test phase, studied old items (i.e., targets) intermixed with new distractors (i.e., lures) are presented one at a time in a source-neutral way. Participants are asked to indicate their decisions of old/new recognition for each item together with the source. These item and source decisions can be collected simultaneously by presenting both source options and the “new” response option at the same time. Here, responses “source A” and “source B” are evaluated as “old”. Alternatively, one can first collect item decisions (old or new), and if the response is “old” for an item, source decision for that item (source A or source B) can be asked consecutively. Put differently, conditional on item recognition, source decisions can be collected in immediate succession (sequential) to item decisions. These standard testing procedures (e.g., Lindsay et al., 1991; Marsh et al., 2006) are illustrated in Figure 3.

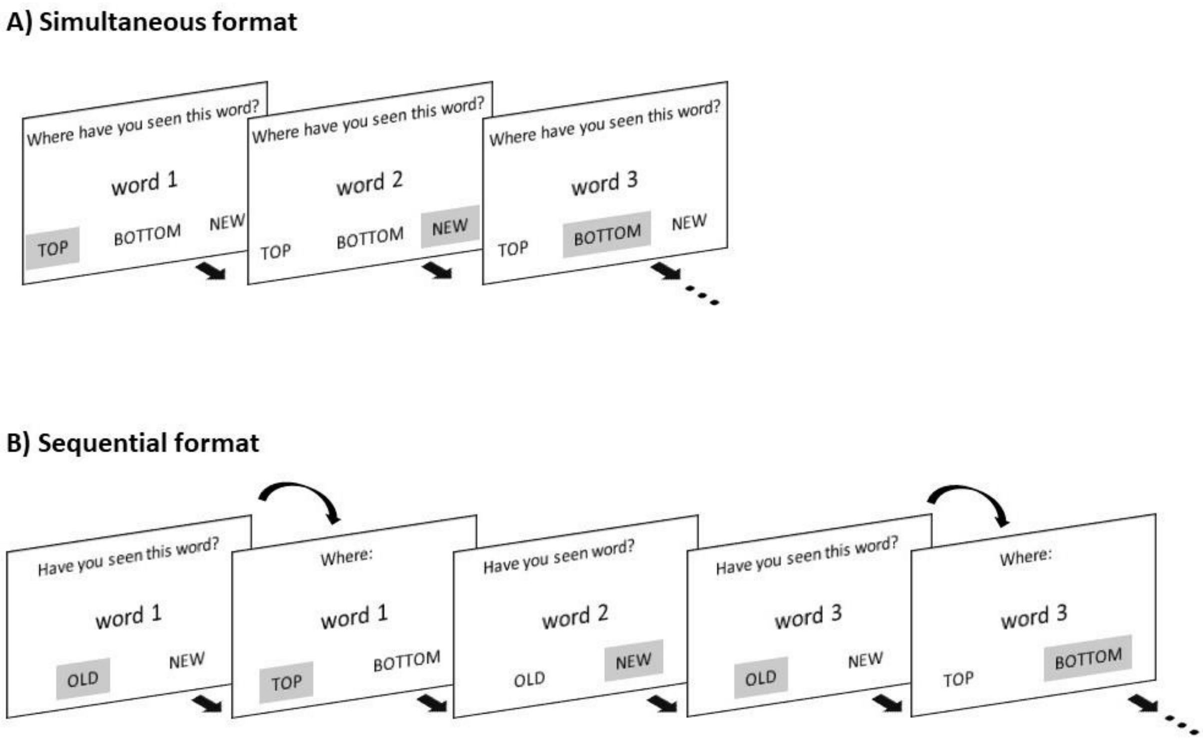


Figure 3. Standard source-monitoring test formats with items (here: words) and sources (here: screen positions [top vs. bottom]). Adapted from Kuhlmann et al. (2021).

The experimental paradigm, regardless of employing the simultaneous or sequential test format, allows researchers to measure both item and source performance. That is, item memory performance can be computed with *corrected recognition scores* (i.e., item hits minus item false alarms; e.g., Chalfonte & Johnson, 1996). However, measuring source performance requires relatively more attention (see Kuhlmann et al., 2021) because it is prone to be confounded by item memory. As a remedy to influences of item memory over source performance, the conditional source identification measure (CSIM) can be calculated for each source and then averaged (ACSIM) for an overall source performance (see Murnane & Bayen, 1996). Nonetheless, the ACSIM score should be carefully interpreted because it cannot account for separate contributions of *memory* and *guessing* processes (cf. Bröder & Meiser, 2007;

Murnane & Bayen, 1996). In order to provide further clarification at this point, it is worthwhile to revisit Johnson et al.'s (1993) source-monitoring definition in the next section.

2.2 Memory Processes and Judgment Involving Guessing

The SMF refers to "the set of processes involved in making attributions about the origins of memories, knowledge, and beliefs" (Johnson et al., 1993, p. 3). Accordingly, source information can be retrieved on the basis of remembering or reconstructed through guessing strategies in cases where remembering is not enough to infer the context (Riefer et al., 1994). Consequently, it is not clear from performance measures (e.g., ACSIM) whether correct responses reflect memory-based processes, guessing-based processes, or some combination of these.

To disentangle the multitude of processes in source monitoring, there are two competing model families, namely, (1) threshold models (e.g., Bayen et al., 1996) and (2) signal-detection models (e.g., DeCarlo, 2003). They differ in characterizing the nature of underlying processes. Specifically, threshold models assume that memory is represented by discrete states, whereas signal-detection models emphasize continuous strength. They also diverge from each other on whether there is source memory for unrecognized items. Threshold models predict that source discrimination is only possible in the state of item recognition. In contrast, this prediction does not hold for the signal-detection model, indicating that source discrimination does not necessarily rely on item recognition. This fundamental difference serves as a testbed for these models (e.g., Malejka & Bröder, 2016; Starns et al., 2008; also see Fox & Osth, 2022, for an overview), but apart from model comparisons, it draws attention to a conceptual perspective: The time-course of item and source processing.

3 The Time-Course of Source Monitoring Processes

"The SMF does not assume an invariant two-stage process in which items are first recognized as old and then attributed to particular sources. It sometimes occurs that an item is initially recognized as old and then attributed to a particular source, but on other occasions an item might first be identified as coming from a particular source (e.g., speaker A) and on that basis experienced as old."

D. S. Lindsay (2008, p. 332),
a pioneer in source monitoring

In the original paper, Johnson et al. (1993) described item recognition and source discrimination within the SMF by outlining their difference under the concept of "differentiation" (see Figure 4). In particular, different memory tasks vary in the degree of differentiation requirements, such that source monitoring needs higher levels of differentiation than old/new recognition. One thus needs more detailed memory to differentiate between different response options during the source test compared to the item test.

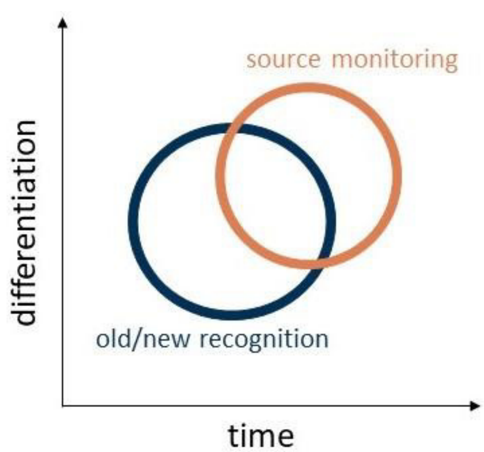


Figure 4. The concept of differentiation. Adapted from Johnson et al. (1993).

As is evident from Figure 4, there is a proposed association between differentiation and time. The question, then, is whether less differentiated information becomes available at earlier stages of processing. Empirical findings based on an unpublished work (Johnson et al., 1983, as cited in Johnson et al., 1994) showed that source decisions were slower than item decisions. However, a critical constraint on this conclusion is that this was a between-subjects experiment, meaning that one group received the old/new recognition test with the binary response options, whereas another group took the simultaneous source-monitoring test and thus decided among more response options (i.e., source A, source B, or new). As a result, this might have somehow confounded RT differences across the different test groups. Following this concern, Johnson et al. (1994) conducted a new experiment focusing on the time-course of reality monitoring. They measured item recognition and source discrimination in the simultaneous source-monitoring test, rendering memory type as a within-subjects factor in the study design. Using the response-signal technique (cf. Reed, 1976), they manipulated retrieval time systematically and instructed participants to make their speeded responses. They measured guessing-corrected item and source memory accuracy with multinomial modeling (see Batchelder & Riefer, 1990) at various response-signal lags. Consistent with the SMF, they observed earlier availability of item memory than source memory. Later on, subsequent reanalysis of Johnson et al.'s (1994) data (Kinjo, 1998; McElree et al., 1999) underlined procedural problems of the original study. However, when the limitations of Johnson et al.'s (1994) experiment were eliminated with a more extensive design (e.g., using more response lags, see Kinjo, 1998, Experiment 1), the results *still* revealed earlier accessibility of item memory compared to source memory.

Another piece of evidence for earlier item memory was provided by Spaniol and Bayen (2002) using an external source-monitoring paradigm. Similarly, they combined the response-signal technique and multinomial modeling (see Bayen et al., 1996), but they focused on the time-courses of item memory and source *guessing* (not source memory). They observed that item memory was also available sooner than source guessing. To conclude, prior research employing this common response-signal technique consistently showed that item memory was accessed *before* both source memory and source guessing.

3.1 Research Gap: Serial or Parallel Processing

The earlier attempts to measure the time-course of item and source information availability in source monitoring heavily depend on the response-signal technique. Thus, our current knowledge is also (mostly) limited to this particular methodology. For serial formulations of item and source processing—for example, as outlined in Lindsay's (2008) quote—the response-signal technique is a well-established approach to adhere. Yet, it still leaves another question unaddressed: *Parallel processing*. In general, the later onset of source information than item information does not necessarily mean that item and source information are retrieved in a strict sequence. Instead, there can be still a sequence between them (i.e., item before source), but at the same time, retrieval courses of item and source can overlap in time (remember the concept of differentiation from Figure 4). Interestingly, previous research has overlooked this possibility, and it is still not known whether source retrieval starts only after item retrieval finishes (Figure 5A) or, rather, emerges in parallel to item retrieval (Figure 5B). The next question now arises as to how one can investigate such nuances in processing.

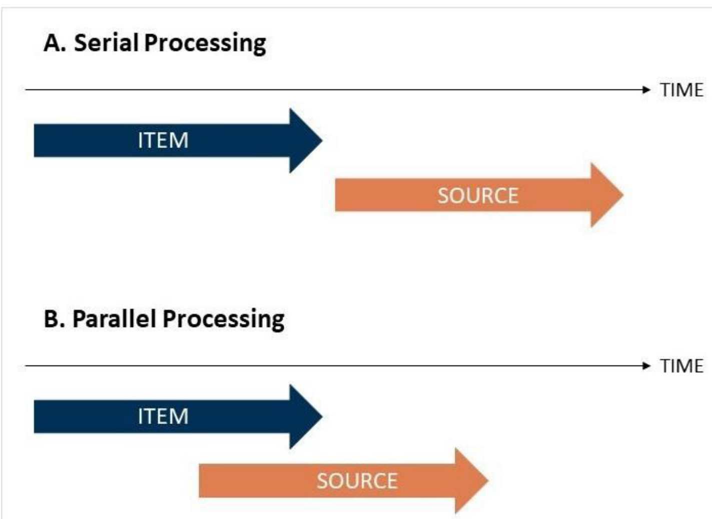


Figure 5. Serial versus (partial) parallel processing of item and source information.

3.2 Reconsideration of Prior Methodology

Despite the feasibility of the response-signal technique and the potential of multinomial modeling, their combination would have three major drawbacks for the current research question (i.e., serial or parallel processing in the source-monitoring paradigm). First, accuracy measured with multinomial-model parameters indicates whether information accumulated sufficiently to pass the threshold. However, this leaves the possibility of parallel item and source retrieval unresolved. That is, it is also likely that source retrieval had already started in parallel to item information but not accumulated sufficiently to exceed the source-discrimination threshold. Thus, this combined methodology only shows whether retrieval is completed at a certain response-signal lag, but it does not explain when retrieval actually starts. Second, based on the core assumption of the multinomial models for source monitoring (for an extensive review, see Erdfelder et al., 2009), source discrimination is only possible in the state of item recognition. Consequently, source memory (and thereby its relevant time-course function) cannot be estimated reliably if item memory is too low under certain circumstances—for example, at very short signal lags (see Spaniol & Bayen, 2002). This should be carefully considered if multinomial modeling is combined with the response-signal technique. Finally, as an inherent aspect of the response-signal technique, the natural course of retrieval is interrupted. This may lead to a confound in retrieval (speed), in a way that the temporal sequence of item and source retrieval may be biased, or even altered, when time pressure is enforced. Given these reasons, I deem it crucial to move beyond the response-signal technique.

Throughout three distinct projects, I pursued alternative routes in order to investigate the serial versus parallel processing in source monitoring and expand the time-course research to *spontaneous retrieval* (i.e., no time restrictions). In the following section, I explain how certain methodological approaches with appropriate designs can contribute to a better understanding of the speed and the temporal aspects of item and source processing.

4 Moving Beyond: Alternative Routes

The main objective of the current dissertation was to test whether we first remember an item and then its source in sequence or whether there is a (partial) temporal overlap between the retrieval of item and source information. On a theoretical level, this dissertation shall serve to fill the important research gap by systematically investigating the mental organization of item and source processing in source monitoring, which will in turn *also* provide an important methodological contribution to the time-course research in general. The following three projects comprise three separate methodologies, respectively: Mouse tracking, the additive-factor method, and diffusion modeling. I summarize each corresponding manuscript in this section. For further details, I recommend consulting the original manuscripts.

Prior to introducing the individual projects, it is important to highlight their common procedural aspects and our reasoning behind them. One commonality across these three projects is the standard sequential test format. The primary goal was to understand item and source processing in the *standard* source-monitoring paradigm. For this aim, we assessed source responses only for items judged to be old (i.e., hits but also false alarms) by employing the (self-paced) standard sequential test format. Remember that this is an experimental setting where item and source decisions can be collected closely in time yet with separable test stages. Moreover, such a design allows both serial and parallel retrieval. On the one hand, item and source can be retrieved in sequence in line with the order they are probed. On the other hand, because participants are tested next for the source if they answer “old” in the item test, they may retrieve source information parallel to item information during the preceding item test as a preparation for the upcoming source test. Therefore, this format is informative about whether there was already some (or even full) source retrieval on the item test.

Another commonality, which was a key manipulation shared across two out of the three projects (see Manuscript 1 and Manuscript 3), is the blocked source-monitoring test format. To gain insight into the processes happening in the standard sequential test format, we *created* a novel blocked design (again, sequential in nature; cf. Osth et al., 2018, for a similar blocked test procedure in source monitoring), serving as the baseline condition. There are similarities and crucial differences between the standard and blocked formats. Similar to the standard format, source decisions in the blocked format are also confined to items judged to be old. However, rather than querying for item and source decisions in immediate succession (as in the standard research of source monitoring), the source test in the blocked format is queried separately. In this source test block, participants are asked to indicate sources of all items they previously judged as old in the order they had responded on the item test. For comparison purposes, in Figure 6, I reiterate the standard sequential test format and introduce our new blocked design.

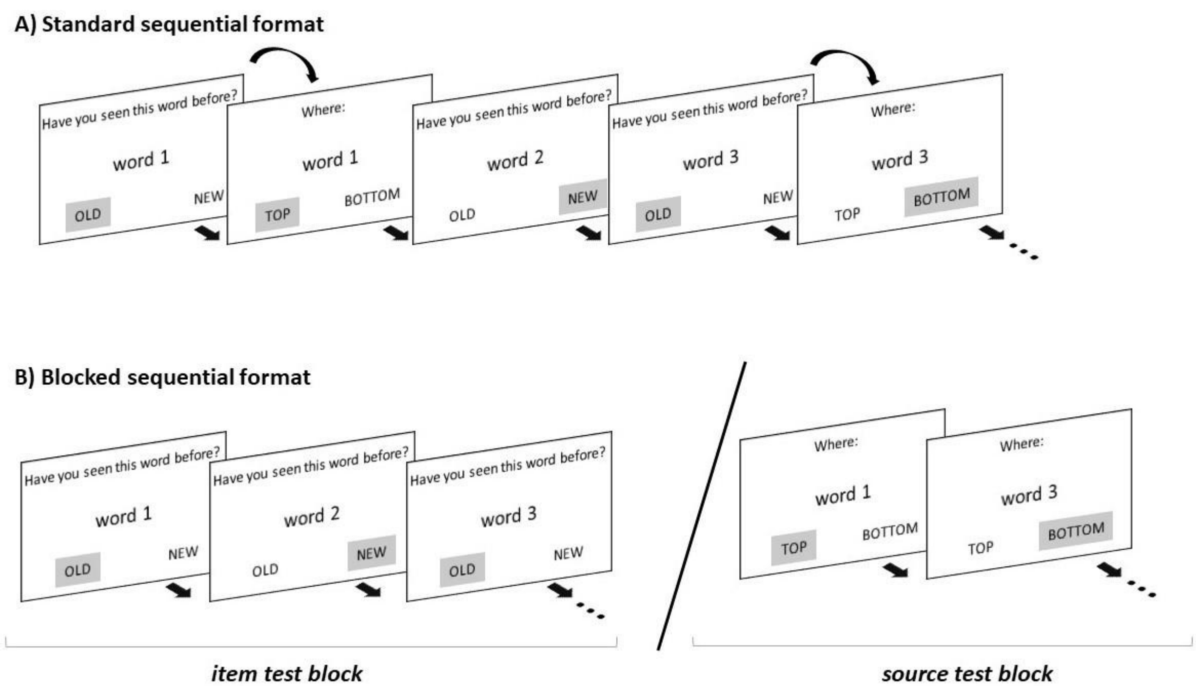


Figure 6. Standard and blocked sequential source-monitoring test formats with items (here: words) and sources (here: screen positions [top vs. bottom]).

We further corroborated the blocked test format with our experiment instructions. Note that item and source learning were always intentional across the three projects, but the prominence of the blocked format is how we informed participants about the ensuing test phase.

More specifically, to improve memory-based responses in the later test, instructions emphasized before study that participants should learn both items and their sources and that they would be informed later which exactly they will be tested on. Next, in the item test block, to maintain the validity of the blocked format, participants were not informed about the upcoming source test block. They were explicitly instructed that only their item memory would be tested at this point, and that source is irrelevant to the responses. We did this intentionally to minimize source retrieval at this stage. In contrast, in the standard format, participants were informed in advance that they would be tested for the source immediately after if they indicated that an item was old in the first step.

In what way can the standard format with reference to the blocked format inform us about the seriality or parallelism of item and source processing? Critically, the merit of the blocked format is twofold. First, the item and source tests are separated in time, hence, they manifest relatively *pure* measures of item and source decisions. Second, source information is less likely to be readily available in the item test (as intended) because participants do not know herein whether—and when—there will be a test for source. In other words, we hamper the efficient parallel retrieval of item and source in the item test block. Comparing patterns of results across the test formats would then be informative regarding the possibility of parallel item and source processing. If item and source decisions are not affected when tested in succession or in a blocked manner, that would favor serial processing. In other words, this would suggest that similar to the blocked format, only the item is processed in the item test stage of the standard format, and subsequently, the source is processed in the source test stage. However, if item and source decisions vary as a function of the test format, that would favor some degree of partial overlap in the standard test and require a detailed investigation of the condition differences (see Manuscript 1 and Manuscript 3).

One could rightfully claim that including a standard simultaneous test (source A, source B, new) might also have provided an interesting baseline condition. However, the simultaneous test format asks for one response only. Roughly speaking, this one response reflects a *sum* of the processes underlying item and source decisions. Therefore, it is not as trivial to infer from the standard simultaneous format, and this is the reason why we did not include it in the current projects although we still believe that comparison to the simultaneous test would be quite enlightening (see General Discussion, for possible future directions).

4.1 Manuscript 1: Mouse Tracking

The SMF predicts that differentiation of item recognition and source discrimination occurs at different rates and develops over time. Mouse tracking indeed provides a promising starting route to measure the dynamics at play here because mouse movements can reflect how item and source decisions unfold qualitatively over time rather than showing the final decision only. In Manuscript 1, we assessed temporal dynamics of item and source memory and tested the question of the seriality versus partial overlap with mouse-trajectory curvatures.

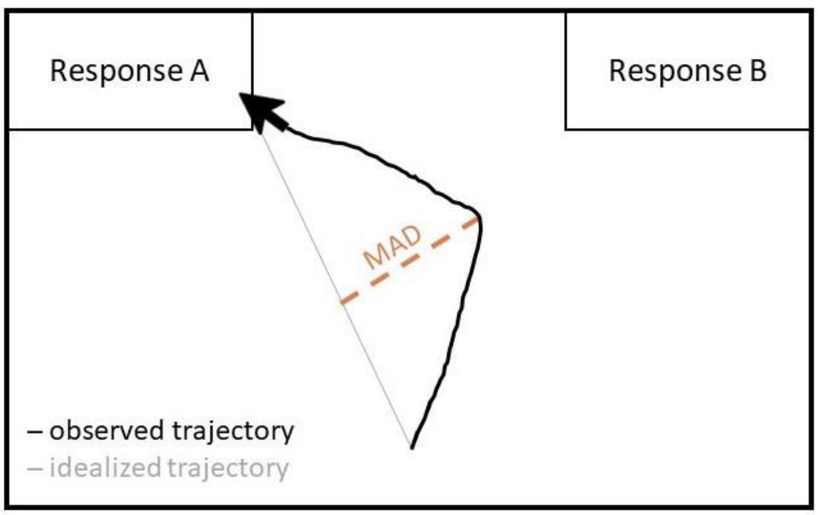


Figure 7. A typical experimental design of the mouse-tracking procedure with one of the frequently used metrics, the MAD (i.e., maximum absolute deviation toward the non-chosen option). Adapted from Kieslich et al. (2020).

In the mouse-tracking procedure, response options are separated on the screen, and participants are asked to indicate their response by clicking on one of the two buttons (see Figure 7). In the meantime, mouse cursor movements are recorded continuously, and different mouse-tracking metrics can be derived therefrom (see Kieslich et al., 2019). Specifically, we focused on the *MAD* values in the current project due to the previous applications of curvature metrics in the old/new recognition tasks (e.g., Gatti et al., 2022) and, more importantly, the link of MAD to source memory (L. Wulff and Scharf, 2020). Note that the MAD stands for the maximum absolute deviation of the observed trajectory from the idealized trajectory, which is an imaginary line between the start and end of the observed trajectory. Therefore, the MAD is an indicator of the (maximum) activation toward the non-chosen response option.

Our design was a 2 × 2 mixed factorial with memory type (item vs. source) as a within-subjects factor and test format (the standard sequential test format vs. the blocked sequential test format) as a between-subjects factor (N = 60). Using variants of the source monitoring test, we investigated how mouse movements during the item and source tests change if item and source decisions are tested consecutively or temporally separated. As reasoned earlier, if we observe no significant interaction between memory type and test format, we would infer a strict temporal sequence. In case of a significant interaction, however, we intended to look at the condition differences closely for the possibility of parallel processing. In this case, our interest was to understand how the relation between item and source memory changes by test format. We expected that this is not driven by a pure change of just item or just source trajectories across the formats per se, but rather the interplay of them.

We first conducted our analyses based on aggregated trajectory curvatures, meaning that the MAD values calculated for each individual trajectory were aggregated first within and then across participants and separately for the standard and blocked formats. The mixed ANOVA analyses using these aggregated MAD values indicated no significant differences across the conditions despite the descriptive trend. Next, to test whether the aggregated trajectory curvatures were representative of the underlying individual trajectories, we performed our linear mixed model analyses with the trial-based MADs as the dependent variable. Importantly, the trial-level analyses revealed a significant interaction between memory type and test format. Follow-up analyses showed that in the blocked format, the source decisions were made less straightforwardly than the item decisions (i.e. the source trajectories were *more* curved than the item trajectories). This was not surprising given higher differentiation and greater recollection demands in source memory (Johnson et al., 1993; Yonelinas, 1999). Critical to our interest, this difference was reversed in the standard format, such that the source decisions were made more straightforwardly than the item decisions (i.e. the source trajectories were *less* curved than the item trajectories).

According to our preregistered hypotheses, findings suggested that source retrieval already started in addition to item retrieval during the item test of the standard format and that this parallel retrieval led to smoother source trajectories in its source test. However, when we probed the interaction in detail, we observed that although the source trajectories were significantly less curved in the standard format compared to the blocked format, the item trajectories did not change considerably between the test formats. This finding yielded another alternative interpretation to parallel processing. That is, source retrieval might have been executed in sequence to item retrieval, but being already in the item recognition state might have rendered source retrieval more accessible in the standard format. Besides, we cannot take it for granted that while participants were working on the source test of the blocked format, item information was totally irrelevant to them. In fact, item information may be a prerequisite for source memory (e.g., Bell et al., 2017), and this may also explain why only the source trajectories differed between the test formats. In sum, the observed patterns during the source tests were not fully conclusive on the possibility of parallel item and source processing, but our results highlighted the close links of item and source retrieval courses.

4.2 Manuscript 2: The Additive-Factor Method

The additive-factor method is an experimental technique built solely on raw (i.e., not transformed) RT data to address questions of the mental organization of processes via selective influences. Considering the close links of item and source retrieval courses, we needed such a sensitive technique to measure these nuanced differences because the additive-factor method was specifically designed to test the strictly serial processing assumption. In Manuscript 2, to the best of my knowledge, we showed the first application of the additive-factor method to source monitoring.

In 1969, Sternberg introduced the additive-factor method as “Extensions of Donders’ Method”. Accordingly, the additive-factor method does not require adding or deleting a cognitive processing stage, unlike the method of subtraction. Instead, the durations of certain stages are lengthened or shortened selectively without affecting the duration of other stages. Therefore, the essence of the additive-factor method lies in experimental factors that selectively change the processing time of the targeted stages. If these factors affect different processing stages selectively, they produce additive effects on mean RT (i.e., significant main effects but no interaction), and this supports a serial arrangement of the corresponding stages. In contrast, if these factors affect a stage jointly, this is manifested by a statistical interaction and supports partially overlapping processes. Put differently, testing an interaction is a means to discriminate between two possible states of affairs: One option is that two processing stages are strictly serial, which is the null hypothesis of additivity. Another option is that a subset of processes operates with some temporal overlap in the same stage, which is the alternative hypothesis because the presence of an interaction violates the assumption that two stages are strictly serial.

To test whether item and source retrieval operate in strict sequence, we had to identify factors that have definite and selective effects on the *speed* of item and source retrieval. Our challenge was that source-monitoring research on latency processes is scarce and that it is not well understood what *selectively* affects the processing speed of item or source. Thus, our next step was to follow previously demonstrated selective effects on memory performance. Intuitively, we expected that better memory would render faster retrieval and that these selective influences on memory performance would ultimately transfer to their retrieval speed. For this aim, we designed a 2 × 2 fully crossed between-subjects factorial design. We manipulated item encoding (generating the study items vs. reading the study items) and source similarity (dissimilar sources vs. similar sources). We hypothesized that generating (compared to reading) would enhance item memory only (Mulligan et al., 2006) and that dissimilar sources (compared to similar sources) would selectively influence source memory (Bayen et al., 1996).

For accuracy analyses, we used the two-high threshold multinomial model of source monitoring (Bayen et al., 1996). To test additivity on the raw RTs, we performed separate ANOVAs for the item and source latency. In Experiment 1 ($N = 128$), source similarity affected source memory performance and source latencies in the expected direction. However, the main effect of item encoding on item memory performance did not extend to item latencies. That hindered us from applying the additive-factor logic further. To understand our unexpected item latency results and to provide another testbed for the additive-factor method, we modified our manipulation in Experiment 2 ($N = 128$) consistent with the encoding specificity principle (see Tulving & Thomson, 1973). Considering the (mis)match between encoding and retrieval for the item context in Experiment 1, we improved the manipulation of item encoding in Experiment 2 without changing the experimental design itself. Experiment 2 showed selective influences on item and source memory performance as well as on their retrieval speed. More specifically, the source effects of Experiment 2 replicated Experiment 1, such that dissimilar sources (compared to similar sources) led to better source memory and faster source retrieval. Furthermore, generating led to faster item retrieval than reading in addition to the memory benefit. Crucial to our preregistered hypotheses, we did not observe a significant interaction between item encoding and source similarity on the item RTs.

Findings revealed that the factors having selective influences on the item and source latencies separately did not interact on the item RTs of the standard format, suggesting that this stage was not occupied by source processing. However, to gain converging support for this *seriality* conclusion, a more detailed examination that depends on both RTs and accuracy is needed. Moreover, other than the *speed* of item and source processing, the latency of such a higher-order cognitive task may include different processes hidden in the current mean RTs.

4.3 Manuscript 3: Diffusion Modeling

The diffusion model (Ratcliff, 1978) is a mathematical model that maps latent cognitive processes in binary decision tasks with separate parameter estimates. By using this model, it is possible to interpret *whether* and *why* task performance differs across conditions at a fine-grained level (Voss et al., 2013). Prior research focusing on recognition memory has shown several applications of the diffusion model (e.g., Ratcliff et al., 2004) and empirical validity of its parameters for the task of interest (Arnold et al., 2015). Spaniol et al. (2006, Experiment 2) further extended its application to a two-choice source monitoring task and observed that the drift rate was sensitive to the source memory processing. Inspired by this study, in Manuscript 3, we implemented the diffusion model for source monitoring and investigated the item and source decision speeds with the drift rate parameter of the diffusion model.

A crucial assumption of the diffusion model is that information accumulates continuously until one of the two response criteria (i.e., thresholds) is reached. The basic model composes four key parameters to explain decisional and nondecisional components that underlie observed responses. *Threshold separation* represents the distance between the response criteria and informs about the amount of information considered in decision-making. The *starting point* of information accumulation shows a priori decision biases. Importantly, the speed of information accumulation is depicted by the *drift rate*. The remaining time outside the decision process such as encoding and motor execution corresponds to *nondecision time*.

Here, we maintained the same design as in Manuscript 1. Thus, our design was a 2 × 2 mixed factorial with memory type (item vs. source) as a within-subjects factor and test format (the standard sequential test format vs. the blocked sequential test format) as a between-subjects factor ($N = 59$). We used the absolute values of the drift rates as a measure of decision speed in each test. Similar to Manuscript 1, if we observe statistically comparable item and source decision speeds across the test formats, we would infer a strict temporal sequence. However, if item and source decision speeds differ by test format, we, again, intended to look at the condition differences closely for the possibility of parallel processing. In this case, we planned to compare the speed of the identical judgments (i.e., item or source) between the test formats.

Based on the trial-level RT and accuracy, we estimated one model per participant, and within each model, we allowed three main diffusion model parameters (drift rate, threshold separation, and nondecision time) to vary as a function of memory type. We performed separate mixed ANOVAs using these individual parameter estimates. Our main results showed a significant interaction between memory type and test format for the drift rates. Follow-up analyses showed that information accumulation in the item test of the standard format was significantly slower than in the item test of the blocked format. This respective difference between the test formats was descriptively (but not statistically) reversed for the source decision speed. This pattern supported our preregistered hypotheses indicating a temporal overlap of item and source processing in the standard format. Our exploratory analyses also showed a significant interaction for threshold separation. More specifically, we observed a larger amount of information considered for an item response in the standard format compared to the blocked format. This corresponding pattern was reversed for the source test, with significantly smaller threshold separations in the standard format. This supported our previous prediction that being already in the item recognition state might require less information to decide on a source response (see Manuscript 1). Additional analysis of nondecision time highlighted the importance of the drift rates by showing the involvement of extradecisional factors in the mean RT differences.

Overall, our test format manipulation affected both decisional and nondecisional aspects of the item and source responses, but the drift rates, which were corrected for these extradecisional factors, allowed us to gain a deeper understanding of processing speed. With respect to our research question, our last project suggested a transfer of source processing to the item test in the standard format, while leaving open details of this temporal overlap (see General Discussion).

5 General Discussion

Maintaining the focus of methodology used in each project, findings are discussed extensively in the respective manuscripts. The primary goal of this section is rather to take a step back and to complete the entire picture from these alternative routes. I acknowledge that there are strengths and limitations in each methodology. However, instead of their superiority to each other, I focus on how they integrate and provide different building blocks by extending our current knowledge, which is mainly based on the response-signal technique. Before discussing our findings from a broader perspective, I provide an overview of the three projects in Figure 8. This summary figure outlines the results and strengthens the connections among the projects.

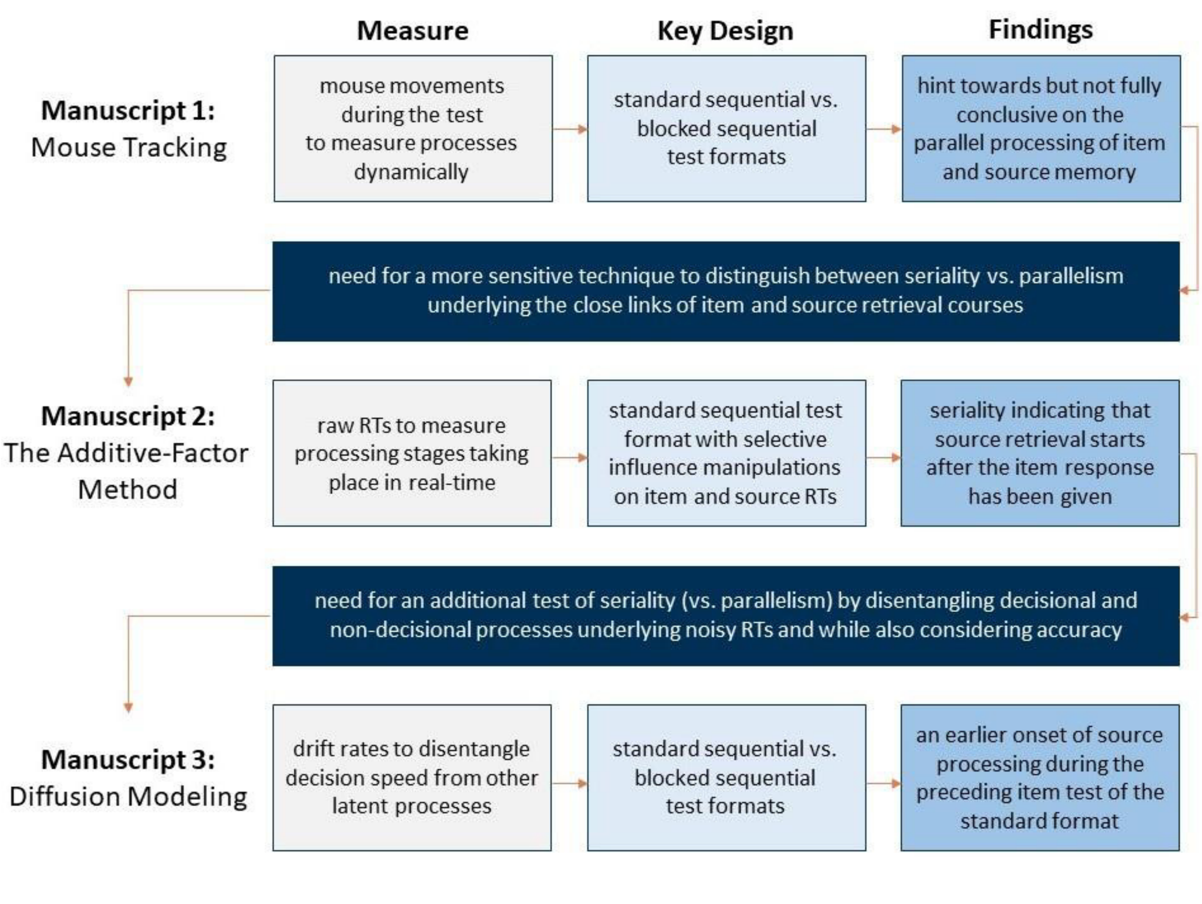


Figure 8. A summary of the present dissertation from the three projects.

5.1 Strengths and Contributions

Albeit with controversy (see Fox & Osth, 2022, for an overview), the possible serial time-courses of item and source memory have been assessed in the literature. However, I am not aware of any source-monitoring research with a central interest in parallel processing. It is crucial to note here that there are indirect postulates of measurement models regarding the possibility of parallelism. For example, multinomial processing tree (MPT) models of source monitoring do not reveal a strictly serial order between item and source memory by also allowing different temporal orders (i.e., fully parallel processes or partially overlapping order) of processes (cf. Batchelder & Riefer, 1999). These are, of course, auxiliary assumptions but do not address the (partially) parallel processing of item and source directly. This dissertation bridges several gaps by benefiting from alternative routes, ranging from traditional RT analyses to recent advances in process-tracing methods.

Mouse tracking is a reasonable starting point because it enables us to track dynamics driving the final decision and thus its total RT (Freeman, 2018). Even for the same responses having the same speed, mouse trajectories can capture their differences qualitatively and present a more detailed investigation of their temporal development. In Manuscript 1, we took advantage of the benefits of mouse tracking and created special circumstances with our blocked design (by separating the item and source tests) to advance our understanding of the standard format. We observed that the pattern observed between the item and source tests of the standard format was reversed in the blocked format. Although these opposite patterns favored the possibility of parallelism occurring in the standard format, detailed analyses following up on this significant interaction showed that differences among variants of the source-monitoring test formats were mainly driven by the source trajectories. In fact, rather than the transfer of source processing to the item test of the standard format, item and source might have been retrieved in sequence, but being already in the state of item recognition might just have eased reaching the state of source attribution. This would also provide an alternative explanation as to why only the source trajectories differed by test format. Consequently, observed trajectory curvatures during the source tests cannot discard the seriality account alone.

In Manuscript 2, we focused on mostly the item part by expecting that if there is any parallel processing happening in the standard format, it must occur before the item response is given. Based on the additive-factor logic, we, again, tested an interaction but this time searched for the additivity of item and source effects on item latencies. We created a sophisticated factorial design by aiming to influence the item and source processing stages selectively. Critically, we observed that the factors that selectively influenced item and source latencies did not interact on the item RTs. With no evidence for their interaction, our results indicated that source effects did not influence the duration of item retrieval and thus favored seriality. Note that we herein tested the null hypothesis of additivity, meaning that we found no convincing evidence for the parallelism of item and source retrieval.

To gain additional support, in Manuscript 3, we repeated our experimental design as in Manuscript 1. As also shown in Donders' method in the first section, there are several latent processes that underlie responses (e.g. motor activity), and this complex nature of decisions might have canceled out or inflated the effects in the raw RTs (in Manuscript 2) or mouse trajectories (in Manuscript 1). We carefully considered these plausible effects and used the diffusion model to disentangle decisional and nondecisional processes underlying item and source responses with separate parameter estimates. Threshold separation, which is specific to the amount of information needed to decide, supported our previous prediction from Manuscript 1 and showed that being already in the item recognition state required less information to decide on a source response. Analysis of nondecisional time further demonstrated that the source-monitoring test formats that we used in the current projects produced additional artifactual effects that are not indicative of processing speed but due to mere extradecisional factors. Most relevant to our research question, the drift rates corrected for these nondecisional processes presented some evidence for the possibility of parallel processing. Specifically, similar to Manuscript 1, we observed a significant interaction indicating different patterns between the test formats, but this time, these differences appeared in the item part such that information accumulation in the item test of the standard format was slower than in the item test of the blocked format. Generally speaking, the source response might have already been prepared in the item test of the standard format, which might have led to a cost in information accumulation in the item test as evidenced by lower drift rates compared to that of the blocked format. This was additionally supported by a larger threshold separation for item decisions in the standard format. Therefore, we concluded that it seems too early to dismiss the possibility of parallelism of item and source processing.

At first glance, our results across these projects can be seen as contradictory. Note that this line of research (i.e., mental organizations in source monitoring) is still in its infancy. Moreover, that does not only pertain to resolving seriality versus parallelism. In particular, Manuscript 2 where we searched for selective influence manipulations on the speed of item and source clearly showed that there is a very limited body of literature about the latency processes underlying the source-monitoring paradigm. Therefore, we are still at the beginning of exploration, and before prioritizing a unified interpretation, we should focus on methodological rigor. A reasonable way to start testing such a new research question is to take advantage of the strength of a certain methodology, determine its constraints, and seek plausible alternatives for those constraints.

The present dissertation introduced different methodologies (mouse tracking, the additive-factor method, and diffusion modeling) as viable approaches to the challenge of seriality versus parallelism and highlighted their ability to compensate for each other's limitations. Mouse tracking (Manuscript 1) provided a detailed look at the temporal properties of data, but given the close links of item and source retrieval, the results remained inconclusive about the parallel versus serial processing. Then, the additive-factor method (Manuscript 2), with its emphasis on factorial interactions and selective influences, favored one of them. It supported seriality by showing the additivity of item and source effects on item latencies. Critically, while interpreting the results based on the additive-factor method, we clearly acknowledged that although our selective manipulations affected overall RTs, it is difficult to ascertain that they exclusively altered the speed of processing. For example, we observed that the source similarity manipulation (i.e., dissimilar vs. similar sources) selectively affected the source (but not item) RTs, but we cannot assure which decisional component(s) was affected. Similar sources (compared to dissimilar sources) might have made participants more conservative or caused a slower information accumulation or both. Therefore, although the additive-factor method is a more sensitive technique to the serial or parallel nature of processing stages than mouse tracking (Manuscript 1) and the diffusion model (Manuscript 3), it cannot resolve this dilemma alone. Analogously, mouse tracking and the diffusion model were not originally designed to test this dilemma but can shed light on the processing with appropriate designs. Lastly, the diffusion model (Manuscript 3) decomposing different processes enlightened alternative explanations that we proposed in our previous manuscripts, and more importantly, revealed some evidence for temporal overlap. Overall, these three projects underlined that the possibility of parallelism can be easily overlooked under the close links of item and source retrieval courses as well as the complex nature of decisions. It is still challenging to explain these different aspects altogether, but different routes facilitate collaborative assessments.

5.2 Limitations and Future Outlook

In 1990, Townsend published an article with a striking title "Serial vs. Parallel Processing: Sometimes They Look Like Tweedledum and Tweedledee but They Can (and Should) Be Distinguished". "Tweedledum and Tweedledee" is a metaphorical phrase describing a pair of things that are highly similar and difficult to be differentiated. Based on Manuscript 3, one can rightfully claim the earlier onset of source processing in the item test of the standard format. However, this temporal overlap also behaves like "Tweedledum and Tweedledee" (see Figure 9). Even though the current dissertation mainly focused on seriality (Figure 9A) and partially overlapping parallel processes (Figure 9B), item and source retrieval can temporally overlap regarding the same test stage but still be executed in strict sequence (Figure 9C). Therefore, our results in Manuscript 3 cannot discriminate details of this temporal overlap.

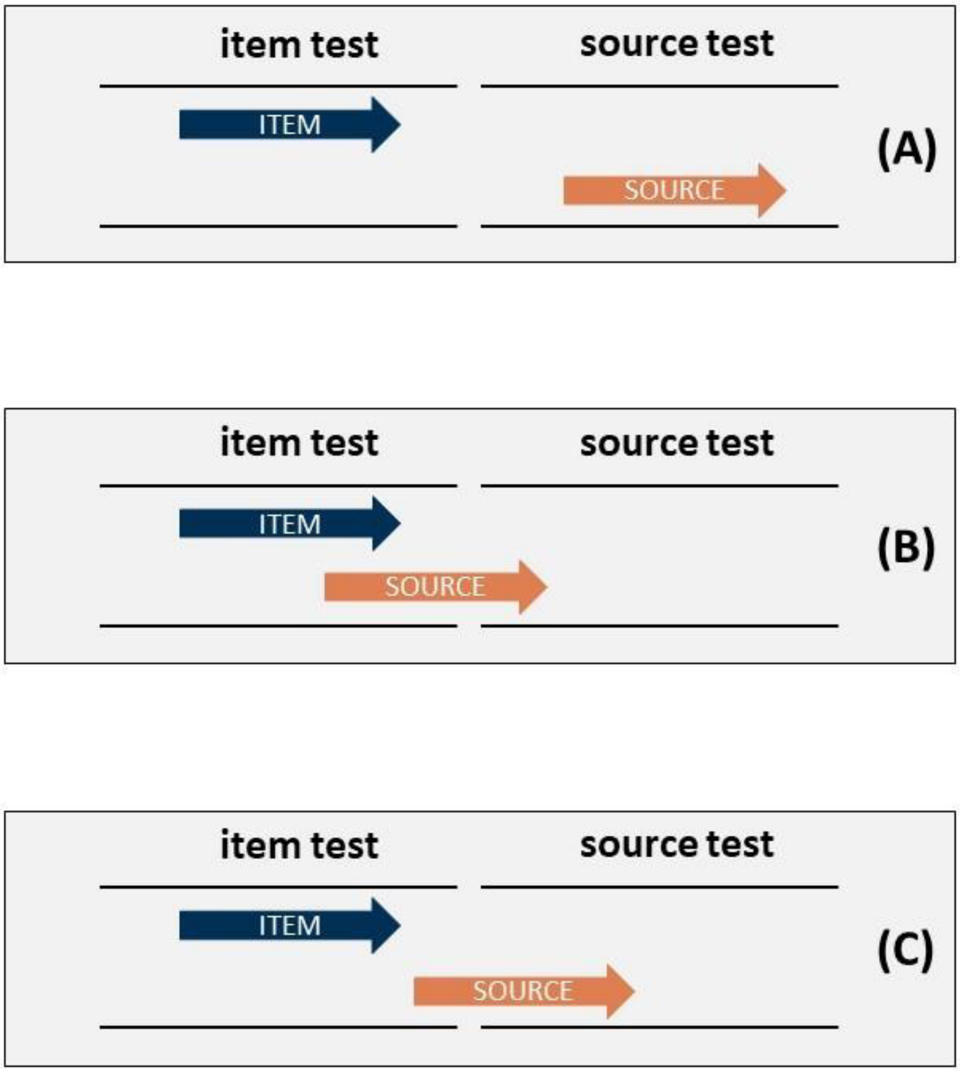


Figure 9. Alternative temporal sequences of item and source.

The time-course question in source monitoring is broader than anticipated. In terms of theoretical curiosity, new avenues considering varying degrees of temporal overlap are desirable for future investigations. However, it is also crucial to remember Lindsay's (2008) quote stating that this is not "an invariant two-stage process" (p. 332). Even if individuals do not engage in parallel retrieval, that does not necessarily mean that they cannot do so at all if instructed to (e.g., using the standard sequential test format with instructions to retrieve item and source simultaneously). Vice versa, the efficient parallel retrieval of item and source might interfere if the source-monitoring paradigm is rendered more effortful (e.g., via dual tasks).

Thus, future research should specify the boundary conditions and investigate under which circumstances the temporal sequence of item and source manifests differently (e.g., for different age groups or via strength manipulations affecting one processing more in magnitude compared to the other).

In the present dissertation, I tested my research question on the standard sequential source-monitoring test format. Consequently, compatible with this standard order of testing, I acknowledge that our results are bound to source decisions for recognized items and that we did not test the reversed temporal sequence (i.e., source before item; also see Malejka & Bröder, 2016). Here, we investigated the possibility of parallelism on the item test part (remember Figure 9). However, above-chance source accuracy for unrecognized items was demonstrated under specific circumstances (e.g., source test also for misses, blocked designs; see Starns et al., 2008, 2013). Therefore, under these designs, it would also be interesting to focus on the possibility of parallelism for the source test part and specifically investigate whether item retrieval emerges in parallel to source retrieval therein.

As a limitation of the present dissertation, although we gain more understanding of our data at the accuracy level by using the two-high threshold multinomial model of source monitoring (2HTSM; Bayen et al., 1996) in all three projects, memory- and guessing-based speed processes that underlie performance are still unclear. Following that, I have been working on a side project (i.e., 2HTSM-RT) that builds on and advances my dissertation. This is a formal modeling approach that integrates RTs with the multitude of processes in source monitoring based on the class of MPT-RT models (see Heck & Erdfelder, 2016, 2020). Figure 10 shows a simple illustration of this extension model. The core idea is to categorize the continuous data into discrete bins from fast to slow responses, separately for each individual (e.g., categorizing responses as “fast” or “slow” relative to the geometric mean per participant). In the 2HTSM model, this can be achieved by adding a latency parameter per branch to estimate the probability of a fast (and complimentary non-fast) response. Importantly, this logic can be applied to both 1-RT settings (i.e., the simultaneous test format) and 2-RT settings (for application to the sequential test format, see Tanyas et al., 2022; for the preregistration protocol, see https://osf.io/gk92x/?view_only=a6c1620b364543738886e777b0076e86). This model allows us to test not only the dependency between speeds of item and source memory but also a wide range of questions on the relative speed of source monitoring processes including guessing, which lie beyond the scope of the present dissertation.

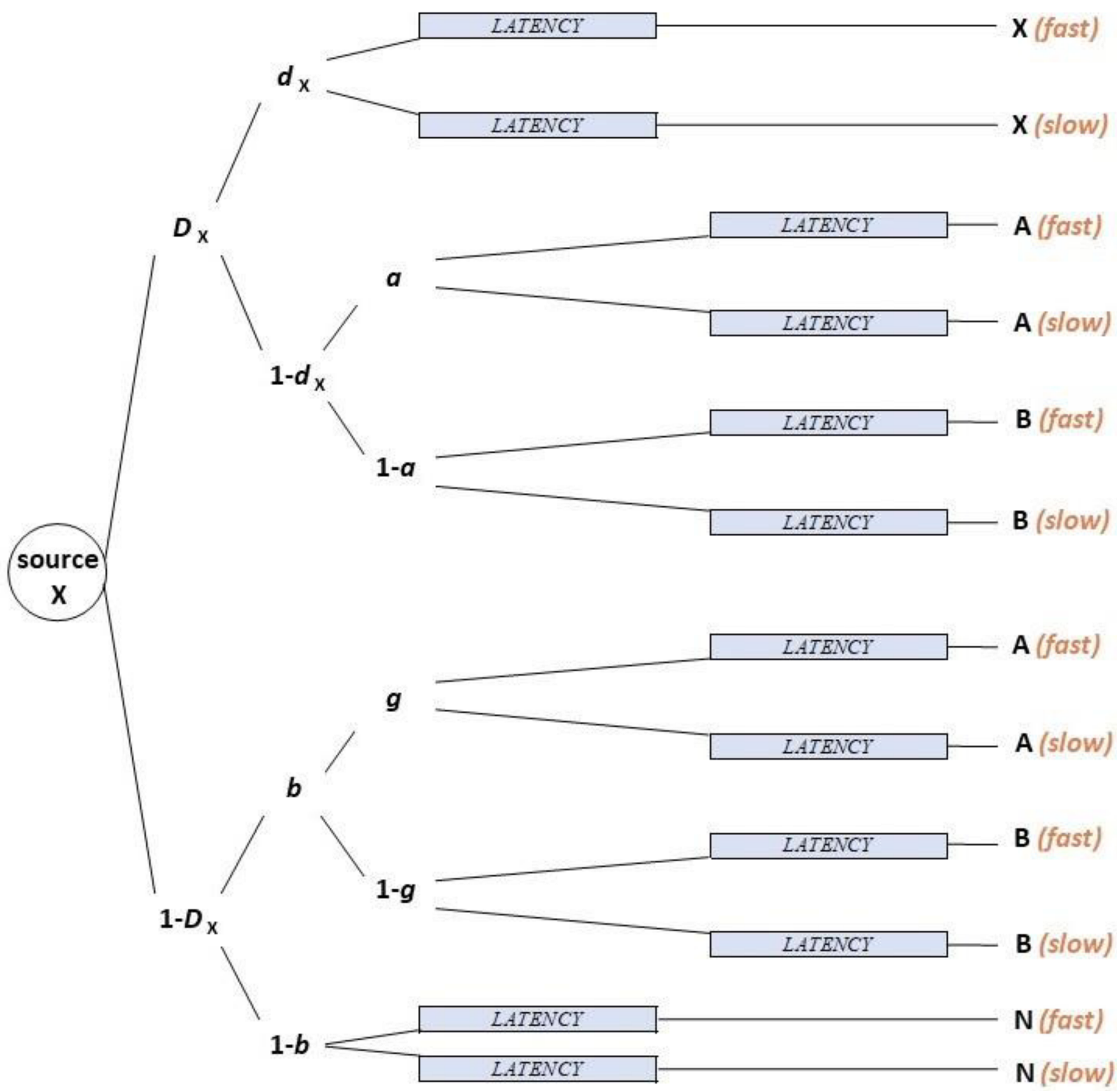


Figure 10. Processing tree representation of the RT-extended 2HTSM model for the 1-RT (simultaneous) test. Labels surrounded by blue squares indicate the new adjustments to the standard 2HTSM model (Bayen et al., 1996). Only one decision tree is presented here. X denotes items from source X, X ∈ {A, B}. “N” refers to “new”. Adapted from Bayen et al. (1996).

The 2HTSM-RT is an ongoing project and might be one of the analytic solutions for a closer look at the latency processes in source monitoring. It is also in line with recent evidence, which indicates a discrete-threshold process for source memory (Zhou et al., 2021; but see Kellen et al., 2021, for item memory). Yet, to broaden our perspectives on source monitoring, alternative suggestions from the same (e.g., RT-MPT models; see Klauer & Kellen, 2018) or different model families are needed. Furthermore, such extensions should not be restricted to RTs only but also draw advantages from other types of continuous variables, like mouse-trajectory curvatures (e.g., Heck et al., 2018).

5.3 Conclusion

Undoubtedly, resolving seriality versus parallelism continues to be a challenge for cognitive psychologists. According to Townsend (1990), this dilemma dates back because “it is inherently related to the capacity of mind and how that capacity is allocated to sundry cognitive and perceptual endeavors.”(p. 46). The present dissertation lacks a clear-cut result for this dilemma but underlined rigorously that the assumption of strict seriality in source monitoring is not warranted. In general, process models appear as the most promising candidates to resolve these fine-tuning differences in temporal sequence.

6 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

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

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old–new recognition. *Emotion*, *17*(1), 120–130. https://doi.org/10.1037/emo0000211

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*(1), 52-60. https://doi.org/10.1027/0044-3409.215.1.52

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*(4), 403–416. https://doi.org/10.3758/BF03200930

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

Donders, F. C. (1969). On the speed of mental processes (W. G. Koster, Trans.). In W. G. Koster (Ed.), *Acta psychologica: Attention and performance II* (pp. 412–431). North-Holland Publishing Company, Amsterdam (Original work published 1868)

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

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

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, *27*(5), 315-323. https://doi.org/10.1177/0963721417746793

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*(59), 1-6. https://doi.org/10.3389/fpsyg.2011.00059

Gatti, D., Marelli, M., Mazzoni, G., Vecchi, T., & Rinaldi, L. (2022). Hands-on false memories: A combined study with distributional semantics and mouse-tracking. *Psychological Research*, 1-14. https://doi.org/10.1007/s00426-022-01710-x

Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review, 23*(5), 1440–1465. https://doi.org/10.3758/s13423-016-1025-6

Heck, D. W., & Erdfelder, E. (2020). Benefits of response time-extended multinomial processing tree models: A reply to Starns (2018). *Psychonomic Bulletin & Review, 27*(3), 571–580. https://doi.org/10.3758/s13423-019-01663-0

Heck, D. W., Erdfelder, E., & Kieslich, P. J. (2018). Generalized processing tree models: Jointly modeling discrete and continuous variables. *Psychometrika*, *83*, 893–918. https://doi.org/10.1007/s11336-018-9622-0

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam: Elsevier.

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

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

Johnson, M. K., & Mitchell, K. J. (2002). Source monitoring. In J. H. Byrne (Ed.), *Learning & memory* (2nd ed., pp. 628–631). USA: Macmillan Reference.

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

Kellen, D., Winiger, S., Dunn, J. C., & Singmann, H. (2021). Testing the foundations of signal detection theory in recognition memory. *Psychological Review*, *128*(6), 1022–1050. https://doi.org/10.1037/rev0000288

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.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, *52*(1), 317–341. https://doi.org/10.3758/s13428-019-01228-y

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

Klauer, K. C., & Kellen, D. (2018). RT-MPTs: Process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, *82*, 111–130. https://doi.org/10.1016/j.jmp.2017.12.003

Kuhlmann, B. G., Symeonidou, N., Tanyas, H., & Wulff, L. (2021). Remembering and reconstructing episodic context: An overview of source monitoring methods and behavioral findings. In K. D. Federmeier & L. Sahakyan (Eds.), *The psychology of learning and motivation* (pp. 79 124). Elsevier. https://doi.org/10.1016/bs.plm.2021.06.002

Lindsay, D. S. (2008). Source monitoring. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 325-348). Elsevier. https://doi.org/10.1016/B978-012370509-9.00175-3

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

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

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

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). The effect of context variability on source memory. *Memory & Cognition*, *34*(8), 1578-1586. https://doi.org/10.3758/BF03195921

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

Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 836-846. https://doi.org/10.1037/0278-7393.32.4.836

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

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). 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

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., 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

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

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 680–693. https://doi.org/10.1037/0278-7393.20.3.680

Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review*, *28*(3), 766-787. https://doi.org/10.3758/s13423-020-01851-3

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

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

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., Pazzaglia, A. M., Rotello, C. M., Hautus, M. J., & Macmillan, N. A. (2013). Unequal-strength source zROC slopes reflect criteria placement and not (necessarily) memory processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1377–1392. https://doi.org/10.1037/a0032328

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., & Erdfelder, E. (2022, March 20-23). *Estimating (guessing-corrected) speed of item and source memory: Extension of two-high-threshold multinomial model of source monitoring* [Conference presentation]. 64. TeaP (Tagung experimentell arbeitender Psychologen; Conference of Experimental Psychologists), Cologne, Germany.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, *28*(4), 363-400. https://doi.org/10.1016/0022-2496(84)90007-5

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*(1), 46-54. https://doi.org/10.1111/j.1467-9280.1990.tb00067.x

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

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

Wulff, D. U., Haslbeck, J. M. B., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting types in movement trajectories. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 131–145). Routledge.

Wulff, L., & Scharf, S. E. (2020). Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us. *Journal of Experimental Social Psychology*, *87*, 103917. https://doi.org/10.1016/j.jesp.2019.103917

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

Zhou, J., Osth, A. F., Lilburn, S. D., & Smith, P. L. (2021). A circular diffusion model of continuous-outcome source memory retrieval: Contrasting continuous and threshold accounts. *Psychonomic Bulletin & Review*, *28*(4), 1112–1130. https://doi.org/10.3758/s13423-020-01862-0

A Acknowledgments

When I approach the end of the 4-year journey, I have come to realize that the final dissertation is the ultimate outcome, but the journey itself is the transformative experience that caused personal and academic growth. This journey was not only filled with accomplishments but also challenges and self-doubt. During difficult times, I always felt to be surrounded by trusted companions. Now, I would like to acknowledge the significance of their presence along the way.

Beatrice G. Kuhlmann. From the very first moment that we met in your office at the Schloss, I sensed a harmony, and this harmony translated into a fruitful collaboration. It was truly a privilege to have been accepted by you to embark on this academic journey under your guidance and mentorship. I sincerely appreciate our long meetings, the always thorough and helpful feedback that I received, and your endless support whenever I needed it. I truly know that you will continue to touch the lives of many students as a mentor, but for me, you will be the *one* that is inspirational in my academic career with your intellectual curiosity and dedication to science. Thank you so much for everything that I learned from you.

Edgar Erdfelder. I deeply admire your passion for learning and your innate gift for teaching. I am grateful for the trust you have placed in my ideas and for providing me with this incredible opportunity to work with you. Thank you so much for your guidance and wisdom.

Arndt Bröder. In our CDSS workshops, your insights and critical evaluation were always helpful and shaped my thinking. Thank you so much for this assistance and for making me honored by accepting to be my thesis reviewer.

Thorsten Meiser. Your breadth of expertise was essential for a comprehensive evaluation of my research. Thank you so much for your kind interest and for the time you spared to assess my dissertation.

Mine Misirlisoy. I am proud to start my academic career with you. Before leaving Ankara, I promised you that I continuously develop myself to the best of my abilities. I hope that I fulfilled the promise that I made to you at the beginning of this journey. Thank you so much for making me feel your support, even from afar.

Pascal Kieslich, *Felix Henninger*, & *Daniel Heck*. Perhaps without your awareness, your achievements broadened my horizons and influenced my research. Thank you for your amazing contributions to the field.

Nikoletta Symeonidou, *Carolin Streitberger*, & *Liliane Wulff*. I was so lucky to have teammates who are also my great friends. You provided me enjoyable and supportive working environment. Most importantly, your kindness and friendship have made me feel welcomed and included. *Christine Emmer* & *Julia Liss*. I always appreciated your positivity and laughter which filled the corridor of L13 with joy and warmth. *Fabienne Unkelbach*. The CDSS brought me a lifelong friendship. I won't forget the experiences we have shared in exploring Manheim together. Thank you, *all*, for making my life in Germany better in every aspect.

My big Turkish family & *my close friends from Istanbul and Ankara*. While I was preparing for my journey to Germany, I packed my suitcase with cherished memories of our time together, placing them at the top. Wherever I go, I carry you all with me in my heart. Your love, support, and friendship have been a constant source of strength and encouragement. *Sevgi Nur*, you offered me the familiar comfort in Heilbronn. My parents, *Fatma Tanyaş* and *Irfan Tanyaş*, thank you so much for your unconditional love.

To my dear spouse, Apo, who has supported me in pursuing my dreams. I dedicate my dissertation to you. Thank you for being my greatest supporter and for helping me navigate the challenges I face, no matter the physical distance between us.

B Statement of Originality

1. I hereby declare that the presented doctoral dissertation with the title *The Speed and the Temporal Aspects of Item and Source Processing in Source Monitoring* is my own work.
2. I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any quotations derived from the works of others.
3. I did not present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.
4. I hereby confirm the accuracy of the declaration above.
5. I am aware of the significance of this declaration and the legal consequences in case of untrue or incomplete statements.

I affirm in lieu of oath that the statements above are, to the best of my knowledge, true and complete.

Signature:

Date:

C Co-Authors' Statements

Co-Author: Beatrice G. Kuhlmann

I hereby confirm that the following manuscripts included in the thesis "*The Speed and the Temporal Aspects of Item and Source Processing in Source Monitoring*" were primarily conceived and written by Hilal Tanyaş, PhD candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences at the University of Mannheim:

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

Tanyas, H., Kuhlmann, B. G., & Erdfelder, E. (2023). *Testing the serial processing model of item and source retrieval: Applying the additive-factor method to source monitoring* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Tanyas, H., Liss, J. V., & Kuhlmann, B. G. (2023). *Information accumulation on the item versus source test of source monitoring: Insights from diffusion modeling* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

I sign this statement to the effect that Hilal Tanyaş is credited as the primary source of the ideas and the main author of all three manuscripts. She derived the research questions, programmed all experiments, collected the data, conducted the data analyses, wrote the first drafts, and was responsible for revising the manuscripts. I refined the theoretical background, contributed to developing experimental methodology related to the research questions, suggested ideas for the analyses and their interpretations, provided recommendations for structuring and improving the manuscripts.

Mannheim, July 2023 Prof. Dr. Beatrice G. Kuhlmann

Co-Author: Edgar Erdfelder

I hereby confirm that the following manuscript included in the thesis "*The Speed and the Temporal Aspects of Item and Source Processing in Source Monitoring*" was primarily conceived and written by Hilal Tanyaş, PhD candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences at the University of Mannheim:

Tanyas, H., Kuhlmann, B. G., & Erdfelder, E. (2023). *Testing the serial processing model of item and source retrieval: Applying the additive-factor method to source monitoring* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

I sign this statement to the effect that Hilal Tanyaş is credited as the primary source of the ideas and the main author of the manuscript. She derived the research question, programmed the experiment, collected the data, conducted the data analyses, wrote the first draft, and was responsible for revising the manuscript. I refined the theoretical background, contributed to developing experimental methodology related to the research question, suggested ideas for the analyses and their interpretations, provided recommendations for structuring and improving the manuscript.

Mannheim, July 2023 Prof. Dr. Edgar Erdfelder

Co-Author: Julia Vanessa Liss

I hereby confirm that the following manuscript included in the thesis "*The Speed and the Temporal Aspects of Item and Source Processing in Source Monitoring*" was primarily conceived and written by Hilal Tanyaş, PhD candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences at the University of Mannheim:

Tanyas, H., Liss, J. V., & Kuhlmann, B. G. (2023). *Information accumulation on the item versus source test of source monitoring: Insights from diffusion modeling* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

I sign this statement to the effect that Hilal Tanyaş is credited as the primary source of the ideas and the main author of the manuscript. She derived the research question, programmed the experiment, collected the data, conducted the supplementary analyses, wrote the first draft, and was responsible for revising the manuscript. I refined the theoretical and methodological background, conducted the main analyses, suggested ideas for the further analyses and their interpretations, provided recommendations for structuring and improving the manuscript.

Mannheim, July 2023 Julia Vanessa Liss , M.Sc

D Copies of Manuscripts

Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-023-02289-z

BRIEF REPORT

The temporal development of memory processes in source monitoring: An investigation with mouse tracking

Hilal Tanyas[1] · Beatrice G. Kuhlmann[1]

Accepted: 31 March 2023

Abstract

The present study investigated whether we first remember an item (e.g., a word itself) and then its source (e.g., position on the screen) or whether the retrieval of item and source information can (partially) overlap. Participants were tested on the source either in immediate sequence to item recognition (as standard in source-monitoring research) or following as a separate block after full completion of the item recognition test to separate these processes in time, providing a baseline. Using the mouse-tracking procedure during the item and source tests, we analyzed how item and source decisions unfolded qualitatively over time. Despite no significant difference in the aggregated trajectory curvatures, more thorough analyses based on the individual trajectories revealed differences across the test formats. In the standard format, trajectories were less curved in the source than in the item test. In contrast, in the blocked format, this difference was in the other direction with source showing more curved trajectories than item. Alternative interpretations of mouse-trajectory curvatures on the source-monitoring paradigm and what their difference may imply for item and source processing are discussed.

Keywords Source monitoring · Source memory · Item memory · Temporal development · Mouse-tracking

Introduction

Source monitoring encompasses memory and judgment processes by which memory records are attributed to their origins (Johnson et al., 1993). Thereby, *source* refers to episodic details that denote the contextual circumstances under which the information itself was acquired. Our focus herein is memory processing in source monitoring, which demands both recognizing the previously encountered items (item memory, e.g., *what* was seen?) and discriminating the origin of those encountered items (source memory, e.g., *where* was it seen?).

Item and source memory are dissociated on a behavioral and neuropsychological level (e.g., Lindsay & Johnson, 1991; Mitchell & Johnson, 2009). However, we do not know yet whether they are also dissociated in time. To date, Johnson et al. (1994) addressed the time-course of *reality monitoring* (a special case of source monitoring, i.e., differentiating internal sources (e.g., imagined events) from external sources (e.g., perceived events)), and found that item recognition was available at earlier response lags than source discrimination. Using a similar response-lag procedure, Spaniol and Bayen (2002) compared the time-courses of item memory and source guessing in the absence of source memory in an external source-monitoring paradigm. However, we are not aware of a study tracking the spontaneous time-courses of item and source memory for external sources. On a theoretical level, Lindsay (2008) speculated about two possible serial time-courses in source monitoring in which either source retrieval may start only after item retrieval finishes, or, alternatively, the source is retrieved first and then provides information for item memory. There is indeed much research and debate on the possible *serial* time-courses of item and source memory (e.g., Bell et al., 2017; Fox & Osth, 2022; Malejka & Bröder, 2016; Starns et al., 2008). Yet, we are not aware of any work querying the possible alternative of *parallel* processing of item and source memory.

The standard source-monitoring test formats either ask for the item and source decision in one step (i.e., Was this item studied in source A, source B, or is it new?) or the source is queried in immediate succession to an "old" response for an item (cf. Marsh et al., 2006). Unpublished response-time

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

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

Published online: 03 May 2023

data from our lab (Tanyas et al., 2022) frequently shows very fast responses on a source query immediately following an "old" judgment, suggesting that participants already retrieved the source during the preceding item query. That is, retrieval of item and source memory may not necessarily occur in a fully sequenced way, despite being probed in that order by the standard testing. Instead, source retrieval may already begin during item retrieval within the same test stage, indicating some degree of "partial overlap."

Mouse-tracking of memory processes

Mouse-tracking is a means to capture continuous neuronal activity in behavior (Spivey & Dale, 2006), and it has become a prominent analytic technique to gain insight into cognition (Freeman, 2018). In this procedure, participants decide between two spatially separated response options on the screen. Meanwhile, their mouse movements are continuously recorded. Tracking cursor positions makes it possible to measure response dynamics in different facets (for an overview of mouse-tracking metrics, see Kieslich et al., 2019).

In recent years, mouse tracking has also been employed in some studies investigating memory via recognition tasks requiring mouse responses. The multifaceted measures of mouse tracking allow researchers to test predictions from different aspects altogether (Gatti et al., 2022) or enable a breakdown of processes subserving recognition. For example, certain metrics of the mouse trajectories can be linked to response bias or encoding strength (Koop & Criss, 2016), while other metrics are related to metacognitive confidence (Papesh & Goldinger, 2012) and inherent memorability (Papesh et al., 2019). Critical to our interest, the pioneering work of L. Wulff and Scharf (2020) implemented mouse tracking to source monitoring and showed that trajectory curvature measured with the MAD (i.e., maximum absolute deviation toward the non-chosen option; Kieslich et al., 2019) is linked to *source memory*. Further, trajectory curvature measured with the maximum deviation from the direct path was also previously assessed in old/new judgments (cf. Gatti et al., 2022). In the following, we thus focus on the MAD considering previous applications of curvature metrics to old/new judgments and, more importantly, its link to source memory.

Overview of the current study

To what extent should memory be detailed to differentiate between alternating response options? As conceptualized by Johnson et al. (1993), it differs by memory tasks, such that source monitoring needs even more differentiation than old-new recognition. Further, they suggest that differentiation of (item and source) memory dynamically changes and develops over time. Here, mouse tracking is a crucial technique to measure such *dynamic* processes, rather than showing only the end-product, by capturing how straighforwardly one opts for a certain response. Thus, we investigated temporal dynamics of item and source memory with mouse movements and specifically assessed trajectory curvature measured with the MAD.

To our knowledge, we are the first study to track item versus source *memory* courses in a standard external source-monitoring paradigm and the first to do so by applying mouse tracking. We manipulated different source-monitoring test formats (the *standard* sequential and *blocked* sequential test) by presenting the source test either in immediate sequence to item recognition (as standard in source-monitoring research) or the source test followed as a separate block after full completion of the item recognition test to separate these processes in time (as our baseline). The blocked format served to provide relatively pure measures of item and source memory, respectively: Even if participants predicted that they will be tested for source at some point, they must not have prepared for it as much during the item test, because the source would only become relevant much later.

We derived separate predictions depending on whether there is a temporal sequence or a (partial) temporal overlap between item and source memory. Intuitively, one would herein expect that the source test would generally create more curvature than its item test because source memory needs more detailed recollection, while recency or non-specified familiarity is sufficient to decide item recognition (Johnson et al., 1997; Yonelinas, 1999). This should particularly show in the blocked format, which more purely measures item versus source retrieval courses, as reasoned above. However, as the direct mapping of mouse trajectories on source monitoring has not yet been explored, we cannot be sure whether this assumed greater required differentiation of source memory (Johnson et al., 1993) indeed translates to more curvature in mouse movements. More crucial to our research question is the comparison of item and source trajectories, regardless of whether they show differential curvatures, between the standard and blocked format:

Hypothesis (H)1. If we observe no significant interaction between memory type and test format, that suggests a strictly serial temporal sequence between item and source memory. That means the difference (or non-difference) between the item and source trajectory curvature is the same and does not matter if tested in succession or in a blocked manner.

H2. In case of a significant interaction, we indicated looking at the patterns of the standard format more closely. If in this format the difference between the source and item trajectory curvature is less pronounced (or even null or in the reverse direction) than in the blocked format, that would speak for a (partial) temporal overlap of item and source memory. Put differently, this would suggest that during the item test of the standard format, participants already began retrieving the source in addition to the item, since they knew they would be tested for source memory following their "old" answer. Consequently, part or all of the curved trajectory shown in the blocked source test was outsourced to the item test in the standard format.

Method

The present study was preregistered in the Open Science Framework (OSF). All materials, including experiment scripts, and results (also supplementary analyses), are available online at https://osf.io/jkrx6/.

Participants

Power analysis using the G*Power-3 software (Faul et al., 2007) for an ANOVA analysis of the aggregate MAD values indicated that a sample size of 60 (i.e., 30 per test format condition) would provide .80 power to detect a medium-sized (i.e., f = .25) within-subjects effect (i.e., of memory type: item vs. source) as well as a medium-sized (i.e., f = .25) interaction between memory type and test format even when conservatively assuming only a .10 correlation between the repeated measures. As these effects were of most interest to our research question, we thus collected data until $n = 30$ was reached for each source-monitoring test format. We acknowledge that our design was only sufficiently powered to detect a large (i.e., $f = .40$) between-subjects effect (i.e., of test format: blocked vs. standard sequential).

Sixty-three German-speaking subjects participated in the experiment. Three participants were excluded from the data analysis because they did not comply with the requirements of the experiment and did not follow the instructions, or else due to technical problems. Analyses were carried out with the remaining 60[1] (43 female, 17 male; M_{age}= 24.92 years, age range = 18–30 years). They were either native Germans (38 participants) or learned German before the age of 6 years (22 participants). The majority (53 participants) indicated a preference for the right hand and all 60 participants reported using a computer mouse with the right hand.

Younger adults were recruited either via the electronic SONA system of the University of Mannheim or via social media groups. We posted our exclusion criteria (i.e., German native or learned German before the age of 6 years; age 18–30 years; no diagnosed/on-going mental health/illness condition) while advertising the study and participants anonymously reported on them in the study. Ten participants were tested in our lab. However, due to the COVID-19 pandemic, we tested the remaining majority of participants remotely if they were willing to install the required software and plug in on their personal computer/laptop under our instructions via video chat. The experimental task lasted approximately 45 min. Participants received either course credit or payment according to our department-set rate of 8€/h. If remote testing took much longer for unforeseen technical issues during installation, we naturally compensated them for the full time.

[1] No participant had fewer hits than false alarms.

Design

The design was a 2 (test format: the blocked sequential test format, the standard sequential test format) × 2 (memory type: item memory, source memory) mixed factorial with memory type as a within-subjects factor and test format as a between-subjects factor.

It is also crucial to note here that spatial position of study words (top vs. bottom) was manipulated within-subjects. Half of these words were presented centered on the top of the computer screen, the other half centered on the bottom. However, as this was preregistered, we did not expect differences in word or position memory between these screen positions and, after ensuring this held in the current data (see Online Supplementary Material), collapsed across this factor in data analysis.

Materials

The item set consisted of 108 emotionally neutral German nouns that were randomly chosen from the Berlin Affective Word List (BAWL-R; Võ et al., 2009) after controlling for certain characteristics (valence: -1.5 to 1.5, arousal: < 3, imageability: > 2, word length: 4–8, number of syllables: 2–3, and frequency: 20–150). From this set, words were randomly assigned to serve as study items (on the top or on the bottom) or distractors for each participant.

Procedure

Automatic stimulus display and data collection were controlled with *OpenSesame* software (Mathôt et al., 2012; version used: legacy backend 3.2.8), using the *mousetrap* plug-in (Kieslich & Henninger, 2017). The experiment was conducted full-screen at a resolution of 1,920 × 1,080 pixels running Windows 10. Remote data collection was limited to individuals whose computer/laptop had the same system qualities and a physical computer mouse (i.e., not touchpad). Thus, these technical features did not differ between the lab and remote testing. The mouse sensitivity settings were left at the system defaults (medium speed, with acceleration enabled). For remote testing, we checked these settings by interacting directly with participants via video chat. Mouse cursor movements were recorded every 10 ms.

Participants were randomly assigned to the experimental conditions upon arrival at the laboratory or recruitment for remote testing. We ensured a comparable distribution across the between-subject groups (i.e., test formats) for lab testing versus remote testing. Before the experiment, participants were requested to complete an informed consent form within the experiment program.

The main experimental task consisted of three phases including a study phase, filler task, and test phase. All stimuli and instructions were printed with 36-point Arial font in black against a white background throughout the experiment. Critically, to increase memory-based test responses, item and source learning were intentional, that is participants were explicitly told before the study phase that they should learn both words and their screen positions, and that they would be informed later which exactly they will be tested on (see below for further details on the instructions). In the actual study phase, 72 German nouns (first letter capitalized in accordance with German spelling) appeared in the upper or lower part of the screen (50% on the top vs. the bottom of screen) for 4 s. A centered fixation dot appeared for 250 ms and a blank screen lasting for 250 ms preceded each stimulus (i.e., 500-ms inter-stimulus interval, in total). Selection of study words, their assignment to the screen positions, and the presentation order were randomized anew for each participant. Participants saw two (fixed) additional primacy buffer items in the study phase that were presented first, one on the top and one on the bottom, and that then along one more (fixed) distractor word served in the practice test.

After the study phase, in order to eliminate the recency effect, participants worked on a 3-min filler task that consisted of basic mathematical equations. Following the filler task, participants were presented with the source-monitoring test, formatted according to their condition. Although they were instructed to respond as quickly and as accurately as possible, all test responses were self-paced. We deemed it crucial that there was no time pressure so that memory processes had ample time to unfold and influence response movements. Before the test session, participants in the standard sequential test condition (cf. Dodson & Johnson, 1993; Marsh et al., 2006; Marsh & Hicks, 1998) were informed that they would be tested for their item memory first, immediately followed by a test for their source memory if they indicated that a word was old in the first step. The 72 old (i.e., 36 top and 36 bottom) and 36 new words were presented in a different random order for each participant. Each test trial began with a start button in the bottom center of the screen (see Hehman et al., 2015; Kieslich & Henninger, 2017). Immediately after clicking on this start button with the computer mouse, a word was shown in the screen center, and the mouse cursor was reset to the exact center of the start button at the bottom center, which enabled us to align each response with an equal starting point. Participants indicated their response as *old* or *new* by clicking on one of the two buttons located in the top-left and top-right corners of the screen (assignment of response options to button location counterbalanced across participants). In this condition, if participants indicated that a word was old, they were next asked to indicate whether it was shown at the top or the bottom of the screen. Similarly, they started this trial of the test by clicking on the start button, and the same word that they just classified as old appeared again in the screen center, with the mouse centered on the start button on the bottom. They indicated their response as either *top* or *bottom* by clicking on one of the two buttons located in the top-left and top-right corners of the screen (assignment again counterbalanced across participants). However, if they responded with *new* in the first item query, the next test trial began immediately. Thus, after they clicked on the start button, a different word appeared in the screen center, and they were again asked to decide whether it was old or new. In the blocked sequential test condition (cf. Fox & Osth, 2022; Osth et al., 2018; Starns et al., 2013), however, before the test session, participants were informed that only their item memory would be tested at this point, and that position is irrelevant for the responses. No mention of the later source test was made to minimize source retrieval at this stage. Thus, in this condition, participants were firstly questioned about whether the words were old or new. The test set-up was exactly the same as in the standard test condition just described, but with the crucial difference that independent of whether *old* or *new* was the given response, no source question was posed (i.e., it immediately proceeded with the next test word as for new responses in the standard test condition). Once participants in the blocked test condition had completed the item test for all words, they were then presented again with all words they previously judged as *old* in the order they had responded and this time asked to indicate their sources, with the same mouse-tracking procedure as in the source test of the standard test. The experimental procedure is illustrated in Fig. 1.

In all tests, participants had to indicate their response by clicking on one of the two buttons located in the top-left and top-right corners of the screen to proceed from each trial. Thus, they needed to answer each trial to complete the experiment, preventing any missing data. Assignment of the response options (old vs. new; top vs. bottom) to the buttons in the top-left versus top-right corner of the screen was counterbalanced across participants. Because counterbalancing was done between participants, the labeling of the response buttons stayed fixed across trials throughout an experiment session to avoid confusion. Participants were additionally informed before the test phase about which option would be presented on which side. Accuracy scores and mouse movements were automatically recorded via the *OpenSesame* scripts. At the end of the experiment,

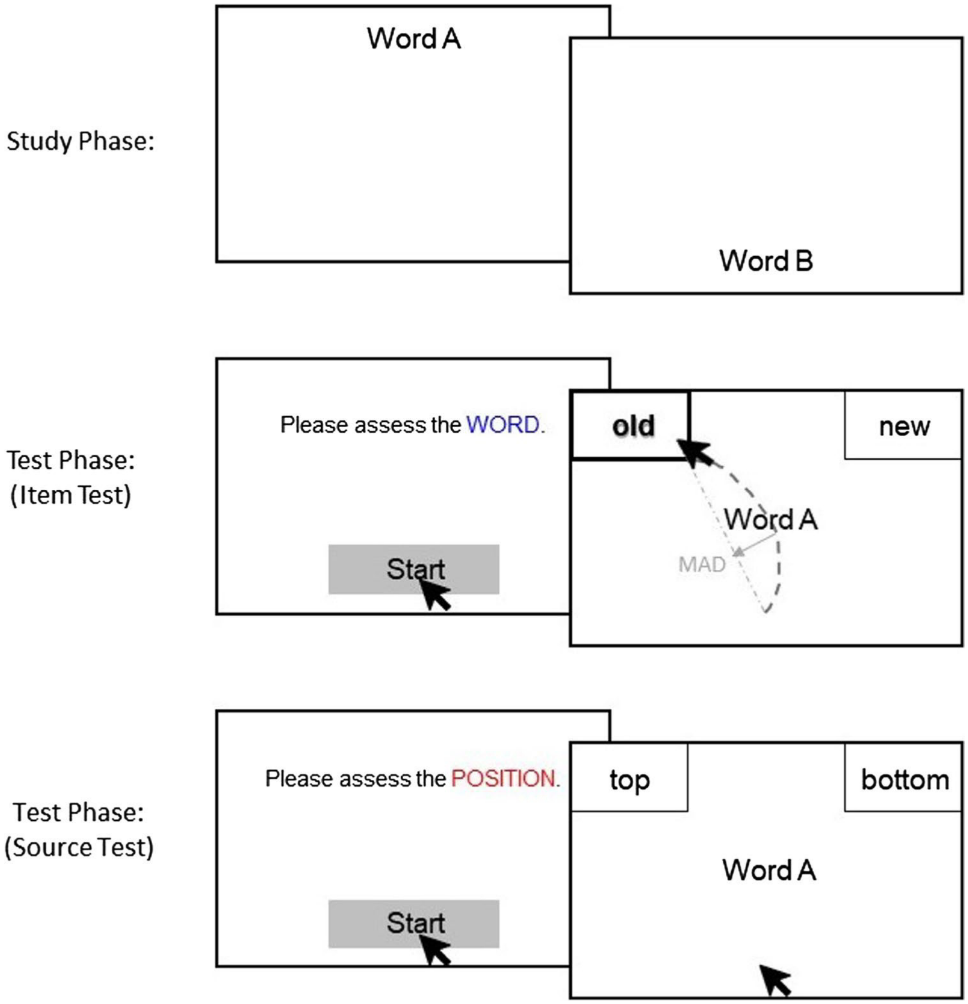


Fig. 1 Mouse-tracking procedure for the source-monitoring paradigm. *Note.* In the study phase, participants saw a number of words (i.e., items) presented either at the top or at the bottom of the screen (i.e., sources). In the test phase, they decided on old/new recognition and source attribution sequentially after a start screen. While participants in the standard format decided item and source decision consecutively for each item upon old response, participants in the blocked format were first asked about their item decision for all items, and then they were asked to indicate the source of the recognized stimuli

participants indicated their demographic information (i.e., age and gender) and indicated their proficiency in German, their handedness and, more specifically, the hand they use for moving the mouse (cf. Kieslich et al., 2020).

Results

We fully followed our pre-registered plan for data preparation and analysis. After reporting the mouse-tracking analyses based on aggregated trajectory curvatures, as planned in our pre-registration, we additionally report more fine-grained analyses based on individual trajectories (cf. D. Wulff et al., 2019; Kieslich et al., 2020). We performed all mouse-tracking analyses in *R* (R Core Team, 2018)[2].

We filtered the mouse-tracking data to analyze only correctly answered trials. Thus, correct source attributions upon correct target detections (41% of targets across both conditions) were included. The total number of accurate trials entering the following aggregated analyses is 933 for the blocked format ($M = 31$ trials per participant, range = 13–54) and 827 for the standard format ($M = 28$ trials per participant, range = 11–50). Information in the Online Supplementary Material additionally shows the multinomial processing tree (MPT) model of source monitoring

[2] Analyses and visualization of the mouse-tracking data relied on the ggplot2 package (Wickham, 2016), the dplyr package (Wickham et al., 2019), the tidyr package (Wickham & Henry, 2019), the afex package (Singmann et al., 2018), and the MBESS package (Kelley, 2017).

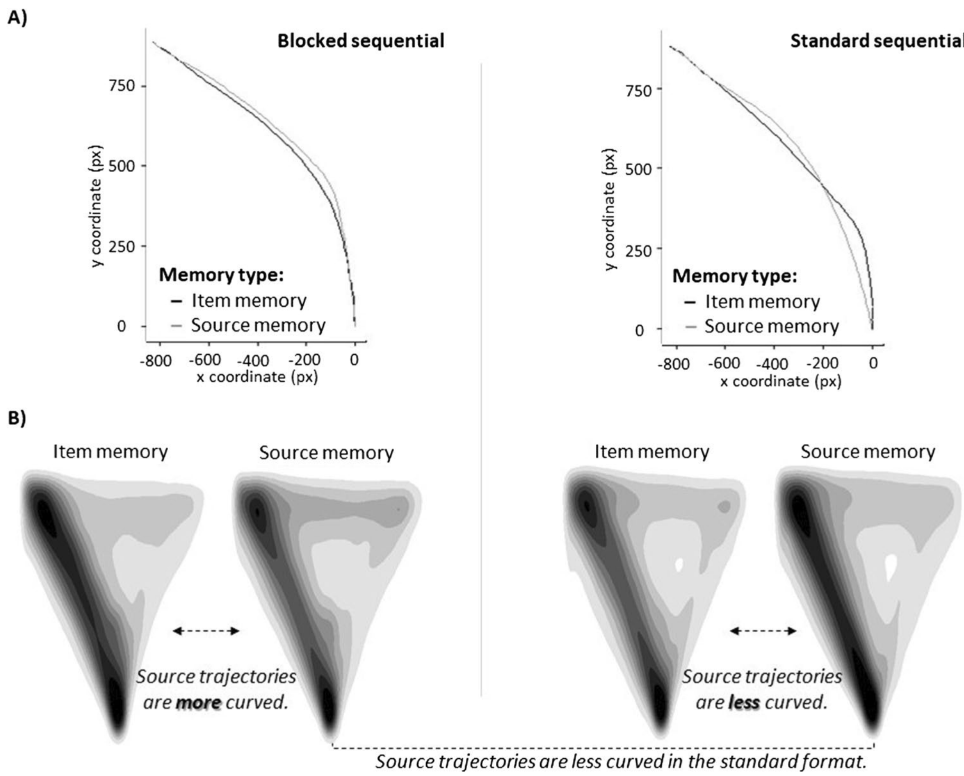


Fig. 2 Aggregate and individual mouse trajectories. *Note.* Left and right panel indicate the mouse trajectories in the blocked and standard test format, respectively. (**A**) Aggregated trajectory curvatures. All answers were flipped to the left and time-normalized. (**B**) Smoothed heat maps of the individual trajectories underlying the aggregate curvatures. This is a graphical illustration for analyzing the trajectories at the trial-level. Darker colors indicate higher density (see also Kieslich et al., 2020). Although the straight trajectories are less common (i.e., trajectories are more curved) in the source test of the blocked format compared to its item test, the reversed pattern is displayed for the standard format in which its source test includes more straight trajectories (i.e., trajectories are less curved) relative to its item test

(Bayen et al., 1996) for the present data as a more fine-grained analysis of the memory processes involved.

Analyses based on aggregated trajectory curvatures

Trajectory measures were derived as follows using the *mousetrap* R package (Kieslich et al., 2016). From the raw data, we extracted the x-y coordinates of the cursor across the interval from the start of the test screen until the response in 10-ms steps (Kieslich et al., 2019). As the correct answer was sometimes to the left and sometimes to the right, we remapped all trajectories to one side. Thus, we flipped all trajectories that ended on the right response option to the left. Of course, given the variation in (self-paced) response times, the total number of recorded coordinates varied across trials. Therefore, we applied the time-normalization function, which divides each trajectory into 101 equally spaced time steps. Then, we computed the MAD for each trajectory (Kieslich & Henninger, 2017).

After preprocessing data, we aggregated the trajectories per memory type, first within and then across participants, and separately for test formats. Figure 2A displays the aggregate trajectories that appear to only differ in details. To test for differences statistically, we conducted a repeated-measures ANOVA using the aggregated MAD values per participant with the within-subjects factor memory type and the between-subjects factor test format. Neither the main effects of memory type, $F < 1$, nor test format, $F(1, 58) = 1.06$, $p = .307$, $\eta_p^2 = .02$, nor their interaction, $F(1, 58) = 2.76$, $p = .102$, $\eta_p^2 = .05$, were significant. However, there was some variation around the mean estimates as well as a descriptive trend capturing that either item or source trajectories were numerically more curved differed by test format (Table 1). We additionally performed a Bayesian repeated-measures ANOVA with JASP (Wagenmakers et al., 2018) and assessed the likelihood of data under one alternative hypothesis relative to the null hypothesis on the basis of Bayes factors (BF_{10}). We report the Bayes Inclusion Factor (BF_{Incl}) across matched models. There was weak-to-moderate evidence for the null hypothesis for the main effects of test format ($BF_{\text{Incl}} = 0.36$) and memory type ($BF_{\text{Incl}} = 0.25$), but the results suggested ambiguous evidence regarding the interaction ($BF_{\text{Incl}} = 1.01$), warranting further analyses based

Table 1 Means (and standard deviations) for aggregated MADs (maximum absolute deviation toward the non-chosen option), and paired t test for the comparison of memory type

Condition	*N*	Item memory	Source memory	*t* test		
				t	*p*	*d*
Blocked sequential	30	289.08 (172.26)	316.30 (162.49)	0.68	.505	0.12
Standard sequential	30	299.02 (195.62)	234.04 (160.30)	-1.70	.099	-0.31

Note: MAD values were aggregated per participant and memory type in each test format condition. More curved (less straight) trajectories are represented by increased MAD values

on the trial-level to test whether our aggregate MAD results were an artifact of condensing the individual trajectories.

Analyses based on individual trajectories

For MAD values, a linear mixed model accounts for intraindividual variation in a more efficient way than the current averaging per person does (cf. L. Wulff & Scharf, 2020). We conducted our linear mixed model analyses[3] with the *lme4* (Bates et al., 2015) and the *lmerTest* R package (Kuznetsova et al., 2017). We included memory type and test format as effect-coded predictors, their interaction as well as a random intercept[4] per participant (Table 2). Critically, the results showed a significant interaction of both predictors, $b = 120.87$, $t(3456.88) = 4.47$, $p < .001$. Next, we compared the model with and without the interaction to verify whether the interaction is needed to explain the data (e.g., Baayen et al., 2008). The likelihood ratio test showed that the model including the interaction explained significantly more variance, $\chi^2(1) = 19.89$, $p < .001$.

To follow up on this interaction, we conducted post hoc pairwise comparisons (*p* values were corrected with the Bonferroni-Holm procedure) using the *emmeans* package (Lenth, 2019). In the standard format, there was a significant difference between the item and source trajectories such that trajectories were less curved in the source test, $t(3456.9) = 3.20$, $p = .008$. In the blocked format, however, this difference was significant in the direction of more curved trajectories in the source test, $t(3456.9) = -3.12$, $p = .009$ (Fig. 2B). While the source trajectories were significantly less curved in the standard format than the blocked format, $t(74.9) = 2.73$, $p = .031$, the item trajectories did not differ significantly across the test formats, $t(74.9) = -0.52$, $p = .733$. Overall, these results demonstrate that in the standard format, trials in the source test led to less curved trajectories relative to its item test, whereas the corresponding difference was in the opposite direction in the blocked format, and that this significant interaction across the conditions seems to be mainly driven by the source trajectories.

Discussion

For comparison purposes, we employed a blocked test format not typically used in source monitoring research (but see Fox & Osth, 2022) to gain insight into item and source memory processes in the commonly used standard source-monitoring test format. Although the aggregated mouse trajectories indicated no significant difference across tests, the trial-level analyses revealed that trajectories were more curved in the source than in the item test of the blocked

Table 2 Linear mixed model with trial-based MADs (maximum absolute deviation toward the non-chosen option) as the dependent variable

Predictors	*b*	*SE*	*Df*	*t*	*p*
Intercept	281.80	17.29	56.35	16.30	< .001
Memory type	-2.59	13.54	3,456.88	-0.19	.849
Test format	40.97	34.57	56.35	1.19	.241
Memory type × test format	120.87	27.07	3,456.88	4.47	< .001

Note: We included the effect-coded predictors memory type (item memory = -.5, source memory = .5) and test format (blocked sequential test format = .5, standard sequential test format = -.5) as well as their interaction. Participants were included as random intercepts. *b* = beta-weight of effect, *SE* = standard error, *df* = degrees of freedom, *t* = *t* values, *p* = *p* values

[3] Although we preregistered that we would explore the individual trajectories, we did not specify this linear mixed model analysis. We thank an anonymous reviewer for suggesting this analysis. It is thus worth noting that the sample size planning was based on our planned analyses at the aggregated -level only.

[4] We also tried a linear mixed model including words as an additional random intercept, but this model was overfitted resulting in a singularity warning (see our R code in the OSF). Thus, we simplified the random structure by removing the intercept of the word (e.g., Gatti et al., 2022). Note that we carefully selected our words as an initial step to control for the noise of items (see Materials section). Further, via the OpenSesame scripts, assignment of the words as targets and lures as well as assignment of targets to the sources were randomized anew across participants, making each participant tied to their own unique random set of items.

format. In the standard format, this difference was reversed, with source showing less curved trajectories than item.

The observed differences confirm the theoretical expectation that the more difficult, recollection-based source memory (with its higher level of differentiation; Johnson et al., 1993) is associated with more curvature than the less difficult, familiarity-based item memory, but only if the source test was delayed from the item test. On the basis of our preregistered hypotheses, this suggests that people may be able to retrieve source information parallel to item information in preparation of the source test in the standard test format. However, we critically discuss this finding and outline open questions as follows. Probing the interaction between memory type and test format further showed that the source trajectories were less curved if tested in immediate sequence to item recognition than tested as a separate block, whereas the item trajectories did not significantly differ by test format. That hinders us from going further merely on the parallelity account and raises another possible explanation of item familiarity serving as a basis for source decision.[5] Specifically, the consecutive testing in the standard format may result in *easier* source retrieval when participants are already in the state of item recognition. Put differently, source processing may not commence during the item test of the standard format (as portrayed by the parallelity account) but rather start with the source query. However, being already in the state of item recognition may just facilitate reaching the state of source attribution. Vice versa, while working on the source test of the blocked format, participants likely did not suppress item information completely, and recognized the item again. This may potentially explain why only the source trajectories differed across the test formats without any costs to the item trajectories. Albeit desirable for further disentanglement in future studies, both of these possibilities suggest close links of item and source retrieval courses, leaving open the challenge of the current research focus. Overall, the time-course question invites a closer investigation of possible patterns of parallelity together with the debate surrounding the serial sequence of item and source memory (e.g., Malejka & Bröder, 2016; Osth et al., 2018).

Mouse-tracking brings a new perspective to this time-course question and provides a useful analytic technique to look at how item and source decisions evolve over time, which is the genuine dynamic process described theoretically by Johnson et al. (1993) under the concept of differentiation. Here, we focused on how straightforwardly participants develop their response in the source-monitoring paradigm as measured by one of the curvature metrics, namely, MAD.[6] Due to their previously demonstrated link to source memory (L. Wulff & Scharf, 2020), we analyzed the MAD values but with a careful consideration of their interpretation. There are varied terms used in the literature describing what trajectories reveal, such as conflict/activation between competing options or one's tentative commitment/attraction to a certain response (Schoemann et al., 2021). For the special case in which L. Wulff and Scharf (2020) investigated stereotype consistency (i.e., consistent vs. inconsistent sources) on source monitoring, the activation of the non-chosen response option can be an indicator of "cognitive conflict." However, in the current study, there is no systematic schema to guide guessing (Bayen et al., 2000) as our aim was to investigate memory processes by simply manipulating the position information, which is regarded as a relatively superficial source cue. Hence, even though the MAD reflects uncertainty in the source monitoring process (L. Wulff & Scharf, 2020), it is as yet unclear whether that is an index of conflict or confidence (cf. Papesh & Goldinger, 2012). Which aspects of mouse trajectories map onto which particular processes depends on the given task (Freeman et al., 2011). As our study seems to be only the second application of mouse tracking to source monitoring, certainly more research is needed.

The present study could guide further research regarding the qualitative nature of memory processing in source monitoring. The results do clearly show that there are pronounced interindividual differences in item and source memory mouse trajectories. Thus, further research should carefully focus on the examination of individual trajectories rather than aggregated trajectory curvatures, as has also been suggested for mouse-tracking analyses in other cognitive paradigms (Kieslich et al., 2019).

Conclusion

Mouse tracking is an insightful way to examine memory processes in source monitoring by exploring the temporal *development* of memory processes over time. Although the evidence is not fully conclusive on the partially overlapping parallel processes of item and source memory, the observed trajectories suggest that querying for item and source memory in immediate succession on a standard source-monitoring task smooths source retrieval compared to when the source is queried in a separate test block. Yet, to draw definite conclusions regarding the possibility of parallel item and source retrieval

[5] We thank an anonymous reviewer for bringing up suggesting this alternative interpretation.

[6] As preregistered, we conducted our analyses on the MAD values. However, interested readers can still find the dataset including the other mouse-tracking metrics as well as response times (RTs) per trial in the OSF.

– especially with regard to the degree of parallel overlap possible – further evidence based on complementary routes from various methodological and analytic techniques is needed.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02289-z.

Acknowledgements We would like to thank Pascal J. Kieslich and Liliane Wulff for their helpful comments on earlier drafts of this work. We benefited from the R codes of Kieslich et al. (2020) and L. Wulff and Scharf (2020), and we are also thankful to them for their open-source materials on the mouse-tracking procedure. Findings were presented at the 62th Annual Meeting of the Psychonomic Society virtually in 2021. We have no conflicts of interest to disclose.

Code availability The *R* code and experiment scripts can be accessed via the OSF at https://osf.io/jkrx6/.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant KU3329/1-1. Hilal Tanyas was further supported by the University of Mannheim Graduate School of Economic and Social Sciences (GESS).

Data Availability The datasets analyzed during the current study are available in the OSF repository and can be accessed via the at https://osf.io/jkrx6/.

Declarations

Conflicts of interest The authors report no conflict of interest.

Ethics approval All procedures performed in the study were in accordance with the principles of the Declaration of Helsinki, the guidelines of the German Psychological Society (DGPs), and the guidelines of the University of Mannheim ethics committee.

Consent to participate Informed consent was obtained from all individual participants included in the study.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic.knowledge about sources in source monitoring. *Memory & Cognition, 28*(3), 480–500. https://doi.org/10.3758/BF03198562

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old-new recognition. *Emotion, 17*(1), 120–130. https://doi.org/10.1037/emo0000211

Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology, 106*(4), 541–551. https://doi.org/10.2307/1422968

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

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science, 27*(5), 315–323. https://doi.org/10.1177/0963721417746793

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2*(59), 1–6. https://doi.org/10.3389/fpsyg.2011.00059

Gatti, D., Marelli, M., Mazzoni, G., Vecchi, T., & Rinaldi, L. (2022). Hands-on false memories: A combined study with distributional semantics and mouse-tracking. *Psychological Research*, 1–14. https://doi.org/10.1007/s00426-022-01710-x

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations, 18*(3), 384–401. https://doi.org/10.1177/1368430214538325

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., & Nolde, S. F. (1997). Electrophysiological brain activity and memory source monitoring. *NeuroReport, 8*(5), 1317–1320. https://doi.org/10.1097/00001756-199703240-00051

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

Kelley, K. (2017). *MBESS: Methods for the behavioral, educational, and social sciences* (version 4.8.1) [computer software]. https://CRAN.R-project.org/package=MBESS

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

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.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods, 52*(1), 317–341. https://doi.org/10.3758/s13428-019-01228-y

Kieslich, P. J., Wulff, D. U., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2016). *Mousetrap: An R package for processing* *and analyzing mouse-tracking data* (version 3.2.0) [computer software]. https://CRAN.R-project.org/package=mousetrap

Koop, G. J., & Criss, A. H. (2016). The response dynamics of recognition memory: Sensitivity and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 671–685. https://doi.org/10.1037/xlm0000202

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. (2019). *Emmeans: Estimated marginal means, aka least-squares means* (version 1.8.2) [computer software]. https://CRAN.R-project.org/package=emmeans

Lindsay, D. S. (2008). Source monitoring. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 325–348). Elsevier. https://doi.org/10.1016/B978-012370509-9.00175-3

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

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

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). The effect of context variability on source memory. *Memory & Cognition, 34*(8), 1578–1586. https://doi.org/10.3758/BF03195921

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1137–1151. https://doi.org/10.1037/0278-7393.24.5.1137

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*(4), 638–677. https://doi.org/10.1037/a0015849

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). 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

Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review, 19*, 906–913. https://doi.org/10.3758/s13423-012-0281-3

Papesh, M. H., Hicks, J. L., & Guevara Pinto, J. D. (2019). Retrieval dynamics of recognition and rejection. *Quarterly Journal of Experimental Psychology, 72*(9), 2328–2341. https://doi.org/10.1177/1747021819836753

R Core Team. (2018). *R: A language and environment for statistical computing* (Version 4.0.2) [Computer software]. https://www.R-project.org/. Accessed Jun 2020.

Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review, 28*(3), 766–787. https://doi.org/10.3758/s13423-020-01851-3

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *Afex: Analysis of factorial experiments* (version 1.0-1) [computer software]. https://CRAN.R-project.org/package=afex

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

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science, 15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

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., Pazzaglia, A. M., Rotello, C. M., Hautus, M. J., & Macmillan, N. A. (2013). Unequal-strength source zROC slopes reflect criteria placement and not (necessarily) memory processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1377–1392. https://doi.org/10.1037/a0032328

Tanyas, H., Kuhlmann, B. G., & Erdfelder, E. (2022) *Does better memory render faster retrieval? Investigating the speed of memory processes in source monitoring* [conference presentation]. 63. Annual meeting of Psychonomic society, Boston, Massachusetts, USA.

Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods, 41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

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

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (version 3.3.5) [computer software]. https://CRAN.R-project.org/package=ggplot2

Wickham H., Francois R., Henry L., & Müller K. (2019). *Dplyr: A grammar of data manipulation* (version 2.1.1) [computer software]. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2019). *Tidyr: Tidy messy data*. (version 1.1.0) [computer software]. https://CRAN.R-project.org/package=tidyr

Wulff, D. U., Haslbeck, J. M. B., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting types in movement trajectories. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 131–145). Routledge.

Wulff, L., & Scharf, S. E. (2020). Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us. *Journal of Experimental Social Psychology, 87*, 103917. https://doi.org/10.1016/j.jesp.2019.103917

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

MPT Model-Based Analyses

The two-high-threshold multinomial processing tree model of source monitoring (2HTSM; Bayen et al., 1996) is advised for obtaining guessing-corrected measures of memory for item and source and separate estimations of guessing bias (for extensive discussion, see also Bröder & Meiser, 2007; Erdfelder et al., 2009). Thus, in the current study, we employed the 2HTSM for disentangling cognitive processes on the basis of response frequencies. Multinomial model fit was assessed via maximum likelihood estimation methods and the G^2 statistic. A p value above .05 indicates that the model explains the data well. Parameter differences were evaluated based on the chi-square difference test statistic ΔG^2 for model comparison. If the p value for the difference test statistic is below .05, the parameter difference is considered significant. Note that we preregistered our MPT model-based analyses but did not indicate any specific hypotheses as the comparisons of memory parameters across the test formats were not the main focus of interest in this mouse-tracking study, so here we rather investigated them as exploratory.

For the current study, the most relevant parameters of the 2HTSM are D and d, standing for item and source memory, respectively. We performed joint MPT model-based analyses to separately measure source-monitoring processes in each condition but in one overarching joint model allowing for comparison. We used the *multiTree* software (Moshagen, 2010) for model fitting and parameter estimation. We followed the identifiable submodels of the 2HTSM (see Bayen et al., 1996, for a detailed overview of alternative model versions). We implemented the most basic and restrictive Submodel 4, as illustrated in Figure S1, with four free parameters (i.e., D [assuming equal detection of items presented on the top or bottom and new distractor items], d [assuming equal probability of remembering the top or bottom source], b, g [assuming equal source guessing when source memory fails independent of item recognition status]) which fit the data, $G^2(4) = 4.46$, $p = .347$. We estimated item and source memory (parameters *D* and *d*; item and source guessing, parameters *b* and *g*, were also estimated but were not of central interest here) across the test formats (see Table S1). After defining our baseline model, we tested the effect of test format by implementing equality restrictions between parameters.

First, we tested the effect of test format on item memory by restricting parameter *D* to be equal across conditions. These restrictions significantly decreased model fit, $\Delta G^2(1) = 12.37$, $p < .001$. Then, we tested the effect of test format on source memory by restricting parameters *d* to be equal across conditions. Again, the model fit became significantly worse, $\Delta G^2(1) = 7.20$, $p = .007$. Overall results suggest that item memory was better in the blocked format while source memory was better in the standard format. In line with Mulligan et al. (2010), item memory was better in the blocked format because the instructions of source monitoring might have changed the sensitivity of old/new recognition in the standard format. Likewise, it is not surprising that source memory was better in the standard format because participants knew they would be tested next for sources upon their “old” answer, and they might have used more stringent criteria at first (cf. Dodson & Johnson, 1993). Next, we tested the effect of test format on item guessing (parameter *b*), and the model fit became significantly worse, $\Delta G^2(1) = 29.31$, $p < .001$. We found a pronounced stronger bias to guess old in the blocked format, which would be expected since participants in the standard format knew they would be tested on source for each “old” response, and thus they only said “old” if they were quite sure. However, participants in the blocked format were not informed about the upcoming source test; consequently, they were more liberal to guess old upon no item detection. Finally, we tested the effect of test format on source guessing (parameter *g*), and we found no significant difference, $\Delta G^2(1) = 0.32$, $p = .574$. However, source guessing averaged across the test formats, $g = .46$, 95% CI [.44, .48], was slightly but significantly below .5,

$\Delta G^2(1) = 11.41$, $p = .001$. That is, there was a bias to guess bottom rather than top, but more importantly, this guessing tendency was comparable across the test formats.

Supplementary References

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

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*(1), 52-60. https://doi.org/10.1027/0044-3409.215.1.52

Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology*, *106*(4), 541-551. https://doi.org/10.2307/1422968

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. *Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42-54. https://doi.org/10.3758/BRM.42.1.42

Mulligan, N. W., Besken, M., & Peterson, D. (2010). Remember-know and source memory instructions can qualitatively change old-new recognition accuracy: The modality-match effect in recognition memory. *Journal of Experimental Psychology, 36*(2), 558-566. https://doi.org/10.1037/a0018408

Table S1

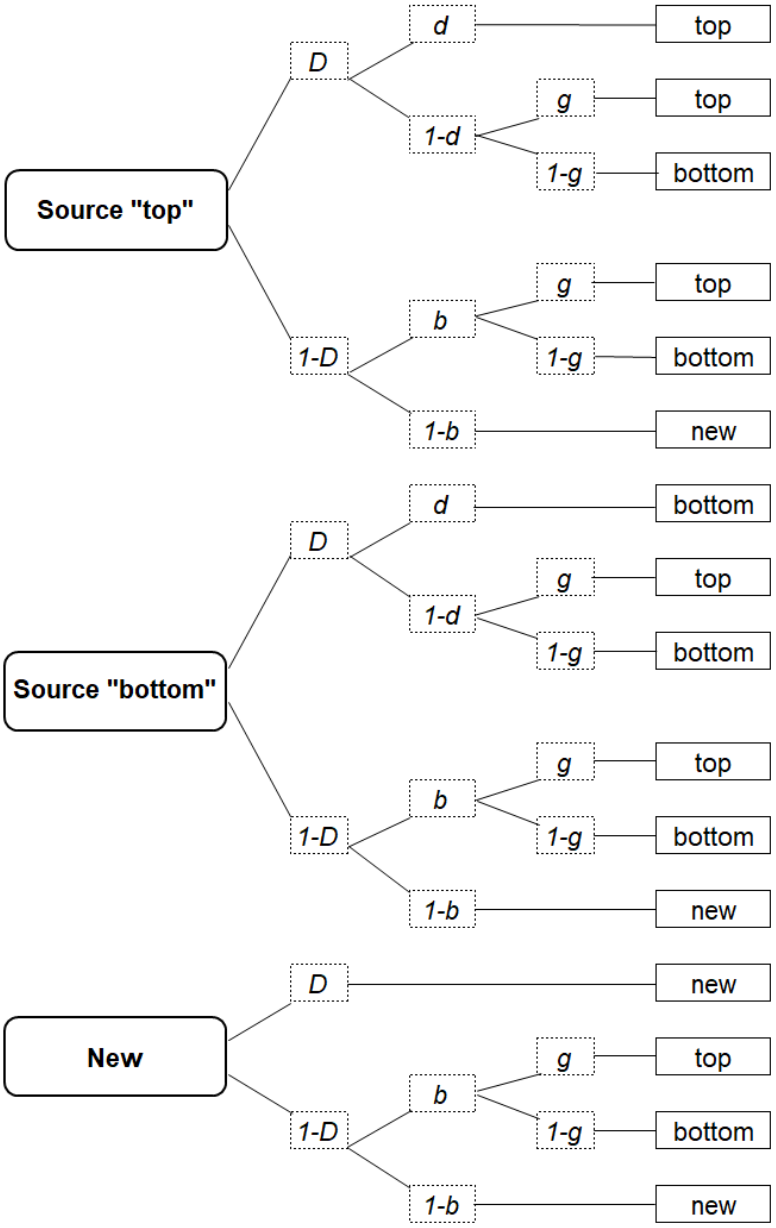
Parameter Estimates and Confidence Intervals for the Four-Parameter Two-High-Threshold MPT Model of Source Monitoring Under Different Conditions of Test Format

	Model parameters			
Test format	*D*	*b*	*d*	*g*
Blocked sequential	.44 [.40, .47]	.42 [.39, .45]	.44 [.36, .52]	.47 [.44, .50]
Standard sequential	.36 [.32, .39]	.30 [.28, .33]	.61 [.51, .70]	.45 [.42, .49]

Note. The presented model parameters are probability estimates that can range from 0 to 1. *D* = item memory; *b* = item guessing (chance level is .5); *d* = source memory; *g* = source guessing (estimates higher than the chance level of .5 indicate guessing bias towards “top”; estimates lower than .5 indicate guessing bias towards “bottom”). Brackets indicate 95% confidence intervals.

Figure S1

Two-High-Threshold MPT Model of Source Monitoring Adapted to Our Source Manipulation



Note. Labels in the leftmost refer to items presented on the source-monitoring test. Labels in the rightmost refer to observed responses. Labels within the branches surrounded by dashes refer to latent cognitive states. *D* = probability of detecting a specific item as old (or a distractor as new); *d* = probability of correctly remembering the source of that item; *g* = probability of guessing that an item was presented by the source “top”; *b* = probability of guessing that an item is old. Adapted from Bayen et al. (1996, Model 4).

Testing the Serial Processing Model of Item and Source Retrieval:

Applying the Additive-Factor Method to Source Monitoring

Hilal Tanyas, Beatrice G. Kuhlmann, and Edgar Erdfelder

Department of Psychology, University of Mannheim

Author Note

Hilal Tanyas https://orcid.org/0000-0002-5849-677X

Beatrice G. Kuhlmann https://orcid.org/0000-0002-3235-5717

Edgar Erdfelder https://orcid.org/0000-0003-1032-3981

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant KU3329/1-1, project number 392884168. Hilal Tanyas was further supported by the University of Mannheim Graduate School of Economic and Social Sciences (GESS). Findings were presented at the 63rd Annual Meeting of the Psychonomic Society, Boston, Massachusetts, USA in 2022 and at the 65th TeaP (Tagung experimentell arbeitender Psychologen; Conference of Experimental Psychologists), Trier, Germany, in 2023. We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Hilal Tanyas, University of Mannheim, School of Social Sciences, Department of Psychology, Mannheim, Germany. Email address: hilal.tanyas@uni-mannheim.de

Abstract

Do item (e.g., *what* was said) and source retrieval (e.g., *who* said it) operate in strict sequence? Sternberg's (1969) additive-factor method is a longstanding technique to test the seriality of latent cognitive processing stages based on individuals' observable response times (RTs). This method searches for selective manipulations affecting the processing time of a single stage without changing the durations of other stages. If experimental factors influence different processing stages selectively, the combined effect of these factors on the mean RT is additive, statistically manifested by significant main effects and no interaction. In contrast, the presence of an interaction conflicts with the assumption that two stages are strictly serial, indicating temporal overlap of subprocesses to some extent. By implementing the additive-factor method to source monitoring, our aim herein was to test whether retrieval processes for an item and its source operate serially or in parallel. Given previously reported selective effects on memory performance, we manipulated item encoding (i.e., generating vs. reading the study items) and source similarity (i.e., dissimilar vs. similar sources) in a fully crossed between-subjects factorial design. In Experiment 1, source similarity affected source latencies, but item generation unexpectedly did not result in faster item retrieval despite the expected memory benefit, preventing further application of the additive-factor method. With a modified procedure, Experiment 2 yielded the expected selective effects on item and source latencies, and the additivity of these effects on item latencies supports the serial model of item and source retrieval.

Keywords: source monitoring, source memory, item memory, additive-factor method, response time

Testing the Serial Processing Model of Item and Source Retrieval:

Applying the Additive-Factor Method to Source Monitoring

Source monitoring subsumes the cognitive processes through which the source of remembered information is determined (Johnson et al., 1993). In daily life, we are constantly engaged in recognizing previous experiences among novel ones, known as *item memory*, along with discriminating their surrounding contexts referred to as *source memory*. The original source of an event (e.g., *Jack* gave me that occupational advice) might be retrieved from memory (e.g., remembering the agent's physical appearance) or derived from judgment processes (e.g., reasoning that the remembered advice is related to Jack's profession and thus likely from him). Similarly, memory and judgment processes contribute to item retrieval. Source monitoring is thus a broad term indicating the whole set of memory (i.e., item and source memory) and guessing processes (on the item and source decision, respectively) while making source attributions (Riefer et al., 1994). While the accuracy-based measurement of these processes and influences thereon have been well studied (e.g., Bayen et al., 1996), little is known about the latencies of the item and source decisions involved in source monitoring. In particular, it is unknown whether item and source retrieval operate in parallel or serially (e.g., first item retrieval, then source retrieval; cf. Lindsay, 2008, for different possible time-courses of item and source retrieval). The objective of the current study was to address this question by applying Sternberg's (1969) additive-factor method to the source-monitoring paradigm.

In the typical source-monitoring paradigm, there is a study phase in which individuals are presented items (e.g., pictures, words) by at least two different sources (e.g., screen positions, agents), and then followed by a memory test on a list of studied (old) items intermixed with unstudied (new) items (cf. Lindsay, 2008). For each item, the standard tests for this paradigm assess old/new and source judgments either in a single joint test (source A, source B, or new?) or sequentially via a separate item test (old or new?) immediately followed by a source test (source A or source B?) to any “old” response (e.g., Lindsay et al., 1991; Marsh et al., 2006).

Previous Research on the Time-Course of Source Monitoring Processes

According to Johnson et al. (1993), source attributions require information to be sufficiently differentiated, more so than necessary for mere item recognition, since correctly recognizing an item as being from one source or another is not enough to avoid source misattributions. In line with higher levels of differentiation required for source monitoring, Johnson et al. (1983, as cited in Johnson et al., 1994), in unpublished work, showed that source decisions take longer than item decisions, as manifested in slower source attribution response times relative to those for old/new decisions. However, Johnson et al. (1994) discussed that this finding may be limited to their specific between-subjects manipulation of the memory test type. In particular, they speculated that the slower response times of the source-monitoring group might have been influenced by the test requiring decisions among more response options (three in total: source A, source B, or new), compared to the binary response options of the test in the old/new recognition group. To address this concern, they specifically focused on reality monitoring (i.e., each picture item was either seen or imagined) and kept response options constant in their 1994 study. They assessed item recognition and source discrimination simultaneously within participants by presenting both source options and the “new” response option at the same time on a single test screen. With the response-signal procedure (cf. Reed, 1976), retrieval time was systematically interrupted during this simultaneous test at four different signal lags. The time-course functions of item and source memory accuracy (estimated with a multinomial model to measure item and source memory uncontaminated by guessing; see Batchelder & Riefer, 1990) for different retrieval times showed earlier availability of item information than source information.

Kinjo (1998) and McElree et al. (1999) reanalyzed Johnson et al.'s (1994) data and revealed uncertainties regarding the time-course of old-new detection and source discrimination. Findings were attributed to noisy data such as using few data points and a low number of response lags. Following this, Kinjo (1998, Experiment 1) replicated Johnson et al. (1994) with an improved design. Critically, the results supported Johnson et al.'s (1994) original conclusion that item information is available sooner than source information.

Notably, the studies by Johnson et al. (1994) and Kinjo (1998) focused on a specific type of source-monitoring called *reality monitoring* (Johnson & Raye, 1981) which refers to distinguishing an internal (i.e., imagined pictures) from an external (i.e., perceived pictures) source. Another prominent type of source-monitoring focuses on discriminating different external sources (e.g., both physically seen in different media or at different locations) that presented information. For this type of source monitoring, we are only aware of one study by Spaniol and Bayen (2002) that investigated the time course with a similar response-signal procedure. However, they did *not* examine the time-course of source memory but rather focused on the time-course of source *guessing* (again, estimated via multinomial modeling) relative to that of item memory. An examination of source memory was not possible because the low item memory at the shortest response lags did not allow for a sufficiently reliable estimation of source memory (see Batchelder & Riefer, 1999). Again, they found that item information became available early on whereas systematic source guessing only emerged at later time lags. However, this does not necessarily have direct implications for the time course of source memory specifically because guessing as an inferential process may emerge later than actual source memory.

Thus, some previous studies have examined the time-course of item and source information availability in source monitoring via the response-signal procedure, but it may not be the best-suited procedure to estimate the time-course of source memory precisely, given its inherent unreliability induced by low item memory for short retrieval times (cf. Spaniol & Bayen, 2002). Further, the procedure puts participants under time pressure and may thus bias the retrieval to occur faster and in a different order than it otherwise would. Finally, the focus on accuracy only shows when item and source information retrieval is completed. It tells us little about when retrieval starts. For example, source information may have been retrieved in parallel to item information at earlier lags but not have accumulated sufficiently to pass the source-discrimination threshold. To examine the *spontaneous* time-course of item and source memory for external sources in a more sensitive and dynamic way, Tanyas and Kuhlmann (2023) recently employed a mouse-tracking procedure and assessed how item and source decisions unfolded over time both in the standard sequential and a blocked sequential source-monitoring tests (i.e., item [old or new] and source decisions [source A or source B] for recognized items were collected either in immediate succession [sequential] or in separate test blocks, respectively). Comparison of the standard way of testing toward the novel blocked format (serving as the baseline condition for "pure" item and source retrieval) indicated that when source decisions were collected in immediate succession to item decisions, the source trajectories were smoother (i.e., decisions were made more straightforwardly) relative to the item trajectories. These smoother source trajectories in the standard sequential test may reflect that participants retrieved source information parallel to item information in preparation for the upcoming source test as they knew that they would be tested next for the source if they answer "old" in the item test. *Alternatively*, however, item and source might have been retrieved sequentially, as how they were queried, but being already in the item recognition state might foster accessibility of source information, resulting in smoother source trajectories. Therefore, even though mouse movements provide a different perspective to the time-course question by showing the temporal development of cognitive processing (see Kieslich et al., 2019), it is, like the response-signal procedure, not the most sensitive technique to distinguish between seriality versus parallelism underlying the close relation of item and source memory.

In summary, previous studies have made use of different techniques to investigate the time-course of item and source information retrieval in source monitoring but the results remain inconclusive as to the seriality versus parallelism of item and source processes. The present research aimed at understanding whether source information (i.e., the contextual details of the information) is retrieved *after* item information (i.e., the information itself) or whether item and source retrieval overlap to some extent. To this end, we employed the widely used sequential test design immediately probing for a source decision after an item is judged "old". Without imposing any time restrictions on this test as in the response-signal procedure or between the item and source test as in the blocked format, this allowed us to investigate the *spontaneous* course of item and source retrieval in a setting where both serial and parallel retrieval are possible but not necessary. We also draw on the advantage that the sequential test provides separate item and source responses, unlike the simultaneous format which asks for one response only. Critically, the sequential order of probing for item first and source second prevents testing the reversed time course of item memory following source memory. However, several previous studies rejected the idea of a reversed time course based on theoretical and empirical reasons (Bell et al., 2017; Malejka & Bröder, 2016; but see Fox & Osth, 2022) and rather favored the established order to measure source-monitoring processes (cf. Johnson et al., 1993). In sum, considering the impact of specific methodological features on item and source retrieval (speed), we herein restricted this general time-course question only to the standard sequential testing of source monitoring, thus assessing source only for items judged to be old. We examined our question through a technique, introduced next, which was specifically designed to test serial versus parallel processing.

The Additive-Factor Method

In cognitive psychology, response time (RT) is extensively used as an indicator of performance and to address the mental organization of processes involved in specific tasks (e.g., McClelland, 1979; Schweickert et al., 2010; Schweickert & Townsend, 1989; Townsend, 1984). There is a general (and reasonable) tendency to assume multiple cognitive stages underlying individuals' observable responses, including both decisional processes that a certain task is designed to measure and non-decisional processes (i.e., test stimulus perception and response execution). Accordingly, while individuals are carrying out a task, they go through different stages from perceiving the stimulus to giving the response (Anderson et al., 2016). Critically, these processing stages may occur either in a sequence of serial steps or in an overlapping manner (Miller, 1993). Serial processing means that information processing proceeds in succession, whereas parallel processing refers to the temporal overlap of subsystems to some extent (Townsend, 1990). The groundbreaking work of Sternberg (1969) aimed to test these alternative mental architectures based on RT data using a technique called the *additive-factor method*.

More specifically, the additive-factor method is a means to test strictly serial processing stages. If one assumes that all processes occur in a non-overlapping sequence, then the observed total RT is the sum of the individual stage durations. The core idea of this method is to identify selective manipulations affecting the processing time of a single stage without changing the durations of other stages. Put differently, one has to determine experimental variables that lengthen or shorten stage durations selectively. These distinct effects by two or more factors are termed selective influence (Townsend, 1984), and this term emerged from Sternberg's (1969) additive-factor logic. It is crucial to note here that, in applications of the additive-factor method, only the duration of a specific stage is prolonged, but that the stage itself remains the same (Townsend, 1990). If experimental factors affect latencies of different processing stages, this implies that they must have additive effects on mean RT if a strictly serial processing model holds. In statistical terms, one should thus observe main effects but no interaction. Such a result would tell us that each stage is occupied by a particular aspect of processing; thus, a set of processes operate in strict sequence. In contrast, if experimental factors affect a stage jointly, this produces a statistical interaction. By implication, both factors are responsible for the duration of that stage and different processes operate in the same stage. In other words, the processes affected by the experimental factors operate with at least some temporal overlap. In the past five decades, various cognitive research areas have employed the longstanding additive-factor method. To our knowledge, it has not yet been applied to episodic memory research specifically, although this has been suggested as a fruitful future venue (cf. Lopes & Garcia, 2014). Inspired by the classical work of Sternberg (1969) and its previous applications, we used the additive-factor method in the source-monitoring paradigm to test whether item and source retrieval is fully sequential or partially overlaps in time, such that source information is already processed when item information processing is still active.

Overview of the Current Experiments

In his seminal study introducing the additive-factor method, Sternberg (1969) investigated the existence of serial processing stages by manipulating the stage durations selectively. Accordingly, one should search for factors having selective main effects on mean RTs of different processing stages. Sternberg (1998) pointed out that “whenever such additive factors are found, and given no stronger arguments to the contrary, it is reasonable to believe that there exists a corresponding pair of stages” (p. 748). We are not aware of any research exploring factors selectively influencing item versus source retrieval *latencies*. However, there is plenty of work demonstrating selective influences on item versus source memory *accuracy* and we deemed it plausible that these may also translate to selective effects on their latencies. For this aim, we followed the rationale of Mulligan et al. (2006) and Bayen et al. (1996) as our benchmark to selectively manipulate the duration of item and source processing. Specifically, Bayen et al. (1996) manipulated source similarity (i.e., sources sharing more features in common vs. sources having diverse features) in a between-subjects design. The results revealed that source similarity had a consistent effect on source memory accuracy but no significant effect on item memory accuracy. Concerning selective effects on item memory, we were inspired by Mulligan et al. (2006). They manipulated item encoding by having a standard read condition (i.e., participants had to read the study items) and a generate condition (i.e., participants had to generate the study items) in a within-subjects design (also studied in a between-subjects design, see Mulligan, 2004). Findings indicated that the generation effect had no reliable influence on memory for the external contexts (here: source memory for location and background color) but rather selectively enhanced item memory. The current route to selectively manipulate item and source memory, and especially their retrieval speed, was the same, hence, our manipulations for item and source processing both occurred at the encoding level.

The standard application of the additive-factor method is on tasks including binary response options. Therefore, the item and source tests of the standard sequential source-monitoring test format each are suitable binary tests (old vs. new; Source A vs. Source B) for application of the additive-factor method. However, we carefully considered its sequential and conditional design (i.e., source test exclusively following items judged to be old). Thus, we only focused on accurate detections for the analysis (i.e., correct source attributions for the correctly recognized items) which always provided both an item RT and a source RTs. In order to test the strictly serial processing assumption, we applied our two-factorial design—item retrieval difficulty (generating vs. reading the study items) crossed with source retrieval difficulty (dissimilar vs. similar sources)—repetitively for the item latency and the source latency. Of interest were the raw mean RTs of the correct responses, that is, RTs were not transformed in any way.

Before conducting our experiments, we preregistered the following hypotheses:

H1. Following Mulligan et al. (2006), we expected that item retrieval difficulty (manipulated by generating vs. reading items at encoding, inducing easy vs. difficult item retrieval later on) selectively influences item memory. Analogously, following Bayen et al (1996), source retrieval difficulty (manipulated by presenting dissimilar vs. similar sources at encoding, fostering easy vs. difficult source retrieval, respectively) was expected to influence source memory selectively. Our intuition was that the selective effects of these difficulty manipulations on memory performance should also extend to processing time: Item retrieval difficulty should affect the item RTs, and source retrieval difficulty should affect the source RTs. We therefore indicated to assess whether these factors influence the accuracy performance as well as the RTs of the specific component as a main effect. This is an assessment of whether the factors have definite main effects on the targeted processes before we can meaningfully interpret the interaction test (cf. Townsend, 1990), with the following exception. Given the possibility that source processing entirely takes place in parallel to item processing, a null effect in source RTs is also possible. Therefore, source retrieval difficulty may not necessarily affect the source RTs, but we can still rightfully test our second hypothesis as long as the selective influence on the item RTs is given.

The crucial test informing about the temporal sequence of item and source retrieval according to the additive-factor method then is the test of the interaction between item retrieval difficulty and source retrieval difficulty on item response time (i.e., using item RT as dependent variable). Depending on the outcome, the result may either support the assumption of strictly serial (H2a) or partially overlapping item and source processing (H2b):

H2a. If the two stages are strictly serial, there should be no interaction between item and source retrieval difficulty on item RTs. Intuitively, one would also expect no main effect of source retrieval difficulty on item RTs. However, if we observe this main effect, as long as there is no interaction, it still speaks for serial processing in separate stages. More specifically, it still indicates that different processes (i.e., item processing and source processing) do not operate in parallel within a stage; rather, such a main effect of the source manipulation on item latency with no interaction would suggest that a separate stage of source processing already commences during the item test after item processing is completed, suggesting separate stages underlying item RT.

H2b. If different processes (i.e., item processing and source processing) operate in the same stage in parallel, there should be an interaction between item and source retrieval difficulty on the item RT. In other words, this pattern is interpreted as evidence that the additivity does not hold, and that source processing already begins in the item processing stage, without waiting for completion of item processing.

Regarding the source RTs, we expected that the time between the item and source response may (not) be affected by source retrieval difficulty as already laid out in H1. Critically, the temporal sequence of item and source retrieval cannot be tested using source RTs following item responses. Since the source test is conditional on the item test, the source RTs might be affected by the preceding item RTs, reflected in an interaction with item retrieval difficulty at this stage even if the processing stages are serial. Vice versa, the absence of item retrieval difficulty effects and interactions at this stage does not necessarily imply strictly serial processing because the item response was already made. Consequently, the primary and crucial test for the interaction effect is on the item RTs (H2a and H2b), because any parallel processing, if it occurs, must occur before the item response is made. The source

RTs were rather investigated to explore the connection of source memory accuracy and source latency.

Experiment 1

The first study was pre-registered in the Open Science Framework (OSF). The preregistration protocol is available online at https://osf.io/p6bnh/?view_only=aa74caa3d87144bb9e121fba470ade2f. All materials, experiment scripts, and results are available online at https://osf.io/9xhsq/?view_only=e72ee9004c914d0db0101565a568c699.

Method

Participants

Power analysis using the G*Power-3 software (Faul et al., 2007) indicated that a sample size of 128 (i.e., 32 per item retrieval difficulty × source retrieval difficulty condition) would provide .80 power to detect a medium-sized (i.e., $f = .25$) interaction between item retrieval difficulty and source retrieval difficulty on the item latency in a between-subjects ANOVA given alpha at .05. In the current study design (i.e., 2 levels per factor), the power for the interaction is the same as the power for each main effect of the same size. Thus, $n = 32$ for each combination of item and source retrieval difficulty also yields sufficiently good power, $1 - \beta = .80$, to detect the medium-sized (i.e., $f = .25$) between-subjects main effects (i.e., the main effects of item retrieval difficulty and source retrieval difficulty on the item or source RTs).

We tested 148 younger adults via the online recruitment platform Prolific (https://www.prolific.co/). The experiment lasted approximately 30 minutes, and participants received payment according to the Prolific-set rate of 6£/hour. We used Prolific's prescreening filters and our self-report demographic survey at the beginning of the experiment to select participants for the following criteria: English as native language (or learned before the age of 6); age (18-30); normal or corrected-to-normal vision; no diagnosis of mild cognitive impairment/dementia; no mental illness with daily impact; no injury to the head that has caused a knock-out for a period of time; no severe respiratory diseases, such as pneumonia or chronic obstructive pulmonary disease (COPD); no medically diagnosed coronary artery or heart issues; no use of medication affecting cognition. Participants received partial compensation for their time spent on the screening survey if they turned out not eligible to continue with the experimental tasks. As preregistered, data from eighteen participants were excluded due to their mean correct RTs for the item or source test being 2.5 or more standard deviations above or below their condition mean (see Results for our reasoning behind this criterion). All participants had the chance to give a general feedback or note any problems in an open text field, if they wished. We also omitted data from two participants who reported that they did not conduct our study carefully. Analyses were carried out with the remaining 128 (99 female, 29 male; M_{age}= 25.19 years, age range = 18 – 30). All participants showed item memory performance above-chance level (i.e., hit rates > false alarm rates). No participant had source memory performance (i.e., ACSIM [average of the conditional source-identification measure, CSIM, for the sources]) that was 2.5 or more standard deviations below the mean performance of their assigned condition.

Design

The design was a 2 (item retrieval difficulty: easy [generating the study items], difficult [reading the study items]) × 2 (source retrieval difficulty: easy [dissimilar sources], difficult [similar sources]) fully crossed between-subjects factorial design (see Shwartz et al, 1977; Yap & Balota, 2007, for other additive-factor studies including factors that were manipulated between-subjects). We assessed both memory types (item and source memory) and their speeds within participants, rendering this a within-subjects factor in the study design. However, we did not analyze it as a factor since we are not interested in this comparison in line with our preregistered hypotheses.

The agents (i.e., sources—who said—) of the statements (i.e., items—what was said—) were manipulated within-subjects. Half of these statements were presented by Jack (as source A), the other half dependent on the source difficulty condition by either Susan or John (as source B). However, we did not expect differences in item or source memory between these sources and thus preregistered our intent to collapse across this factor in data analysis. Likewise, the theme (expected doctor vs. expected lawyer) of the item statements (see Material) was manipulated within-subjects but each source presented *equally* many statements of each theme. Participants were also explicitly informed about the equal likelihood of schema-expected statements presented by each source. Importantly, we did not assign a profession to either of the two sources throughout the experiment. As a result, ideally, participants should not develop a profession-based bias which would be anyways controlled in the model-based analyses. Furthermore, expected doctor and expected lawyer statements can be analyzed jointly because item memory for the two types of schematic statement has been found to be comparable (cf. Kuhlmann et al., 2012). Thus, the theme was irrelevant for our current research interest and only a technical factor in the design in order to obtain enough study trials because the doctor-lawyer paradigm[1] does not provide enough equally expected statements.

Materials

Sources. Two faces (and their names) presented on the computer screen −Jack as Source A and Susan/John as Source B− were manipulated as the sources (see Figure 1). These were black-and-white pictures used by Bayen et al. (1996). Participants either studied the statements spoken by the distinct agents (i.e., via the presentation of one male and one female face and their distinct names—Jack or Susan) or by the similar agents (i.e., via the presentation of the two similar male faces and their similar names—Jack or John).

Items. We used the sentence material of the doctor-lawyer paradigm (Bayen et al., 2000, Experiment 2) as items which has been extensively tested in source monitoring (e.g., Bayen & Kuhlmann, 2011; Hicks & Cockman, 2003). Thus, the item set consisted of "expected doctor" and "expected lawyer" statements. We selected 20 expected doctor (e.g., "Are you taking any other medicine?") and 20 expected lawyer (e.g., "I ask the jury to acquit this man.") sentence pairs, each with a target and distractor version differing only in one word or phrase (e.g., the distractor versions of the previous two examples are "Are you taking any other *prescriptions*?" and "I ask the jury to *hang* this man."). The set of 20 statements from each theme was further divided into two subsets of 10 to serve as items from two different sources in the study phase: One subset (randomly determined) of each theme was presented by Jack and the remaining one was presented by Susan or John. That is, both sources presented equally many statements expected for a doctor and expected for a lawyer. All the subsets were created by controlling for certain characteristics and matched for sentence length (number of characters and words) and expectancy ratings. In the instructions before the study phase, another set of two neutral statements from this pool served as examples and for practice with marking the agent of the statements to familiarize participants with the task and duration of each statement presentation in the ensuing task. An additional set of four sentence pairs (2 expected doctor and 2 expected lawyer) was used for the primacy buffer items (later serving as practice trials in the source-monitoring test). In the test phase, participants were tested on 80 statements including all 40 studied statements (the half had been presented by Jack, and the other half had been presented by Susan or John) and 40 new statements which were the distracting versions of the studied target statements from Bayen et al. (2000).

Importantly, in the generate condition, we used word fragments (i.e., some letters of these words were missing[2] and replaced by dashes). Participants were expected to generate them in accordance with the statement in which they were used (e.g., "Are you taking any other m_d_ _in_?" and "I ask the jury to a_q _ it this man."). The critical word in each statement used as the word fragment in the study phase was the one that differs between the target and the distractor version of the same sentence (see Bayen et al., 2000, for the sentence pairs). In the test phase, it was thus clear to all participants that the test includes distractors very similar to the targets, but for the participants assigned to the generate condition, it would be easier to discriminate as they initially encoded the critical word through deeper processing.

Procedure

The experiment was programmed using the lab.js experiment builder at https://lab.js.org/, which is a browser-based environment (based on HTML and JavaScript) and also allows precise timing performance (Henninger et al., 2022). This online experiment was hosted on the server application OpenLab (https://open-lab.online/; Shevchenko, 2022) with access offered to participants via the platform Prolific (https://www.prolific.co/; see also Palan & Schitter, 2018). Participants needed a PC or laptop to work on the study (completing the study with a smartphone or tablet was technically not possible because participants had to use a physical keyboard to advance throughout the study). Participants received a detailed description of the study and its requirements on Prolific. If they decided to participate in the study, they were redirected to OpenLab to conduct the actual experimental task. They were then randomly assigned to one of the four conditions by using OpenLab's urn function for between-subjects randomization ensuring roughly comparable group sizes. Before working on the actual task, participants were presented an informed consent form within the experiment program. After that, participants first answered demographic and health questions. If based on these responses they did not meet our eligibility criteria, the program terminated and informed them how to receive partial reimbursement for the time spent. Only eligible participants were able to continue with the experimental task.

The experiment session consisted of a study phase, a filler task, and a test phase. Before the study phase, all participants were informed that they would see the faces of two people, Jack or Susan/John, accompanied by a statement spoken by them. All participants were explicitly asked to memorize both the statements (what was said) and their agents (who said it) for the following memory test. Further, the generate and read conditions were instructed on their specific item-processing task (see below). As part of the instructions, they saw two example screens, sampled from the equally expected doctor-lawyer statements but not part of the statements used in the source-monitoring test, to give them a better idea of the task and the encoding condition that they were assigned. Participants saw four (fixed) additional primacy buffer items in the study phase that were presented first, and each source presented two of these (order randomized anew for each participant), and then they later served as practice trials in the source-monitoring test. Following this, 40 statements appeared centered at the bottom of the computer screen in single quotes. Above the sentence appeared the picture of the source in the middle of the screen. Below the picture, centered on the screen, appeared the name of the source in capital letters. With this arrangement, participants were able to focus on all (both the item and the source) of it once. The statements were printed with 24-point (corresponding to 32 px), and the source labels were printed with 45-point (corresponding to 60 px) Arial font size in black against a white background throughout the experiment. The statements and their source (the face pictures labeled with their names) were presented in a different random order for each participant and remained in their respective position on the screen for 6 seconds per trial.

In this *study screen*, the agents of the statements were Jack and Susan (i.e., dissimilar sources) for the easy source retrieval condition and Jack and John (i.e., similar sources) for the difficult source retrieval condition. We presented the same statements across the conditions with the crucial difference that participants were encouraged to encode the items (statements) differently. More specifically, the participants assigned to the easy item retrieval condition saw statements each containing one fragment, some letters of a word were missing. They were informed that the number of dashes is always equal to the number of missing letters and expected to generate this missing word (i.e., the generate condition). However, the participants assigned to the difficult item retrieval condition saw statements each containing one underlined word, and mere reading was expected (i.e., the read condition). In the next *word-entry screen*, the instruction “Please type the word that you generated.” appeared for the generate condition. Here, they were asked to generate the word fitting in the fragment from the previous screen into the input field. However, the participants in the read condition saw the instruction “Please type the underlined word.”. Here, they were asked to rewrite the underlined word from the previous screen into the input field. Participants from both conditions had 5 seconds to give a response (i.e., typing the word and submitting their answer by clicking on the submit button or pressing “Enter”). The cursor always appeared within the input field to save time for typing. If they needed less than 5 seconds, their answer was framed in green to indicate their answer was logged for the remaining time. The next screen was always presented after the 5 seconds had elapsed. Following this, participants from both conditions saw the same sentence printed in full (i.e., with the fragment solved in the generate condition) for 3 seconds in the *feedback screen*, and the critical word was underlined. Our reasoning here was twofold. First, even if the participants in the generate condition failed at generating the critical word, they would have the chance to encode the correct answer and still benefit from trying to solve the word fragments. Thus, they were able to ascertain the correct solution from the feedback (cf. Jacoby, 1991, for a similar design). Second, as a natural consequence of the item encoding tasks (i.e., generating or rewriting the critical words) in the study screen, we may draw participants' attention from source processing (the agents) and direct them to the statements too much (see Mulligan et al., 2006, for a similar argument about the encoding tradeoff). We thus wanted to give participants another chance to encode the sources without the additional need to do the respective item encoding task. Therefore, after the feedback screen, all participants were asked to indicate the agent of the statement (i.e., who said) in the *agent-click screen*. Here, they had 3 seconds to give a response (i.e., clicking on the source picture accompanied by the source label). If they needed less than 3 seconds for their answer, their choice was framed in green to indicate their answer was logged. Then, the next trial started immediately, signaled by a fixation cross in a silver-grey screen lasting for 500 ms. We emphasize here that by using a fixed timing on each screen, we equated the total processing time of the study phase across the conditions. We set the time limits based in prior piloting such that a response could be made without overly rushing but also without providing too ample to allow for rehearsal or to become boring/aversive to participants.

After the study phase, participants judged basic mathematical equations whether each was solved correctly or incorrectly, serving as a 3-minute filler task activity. Following the filler task, participants took the source-monitoring test. Before the test session, participants were informed that they would be tested for their item memory first, followed by a test for their source memory if they answered old in the first step. They were further warned that the new statements would be very similar to the studied statements, differing in one word or phrase only. The test trials were preceded by four (fixed) practice trials (2 targets and 2 lures) in which the primacy buffer items were used (again, order randomized anew for each participant). In the actual memory test, participants saw 80 statements one at a time, and this time all printed at the top of the computer screen and in the complete form (e.g., no words fragmented or underlined). Forty of the 80 test statements were the target items (as presented in the study phase) and therefore belonged to the list of statements participants were asked to remember (called old). Of these target test statements, 20 were expected-doctor and 20 were expected-lawyer sentences. Of each group of 20 items, 10 had been presented by Jack, and 10 by Susan or John. The other half of the statements was similar to the sentences that we had shown participants at the beginning but new (i.e., as stated earlier, we made tiny modifications in their wording, changing only one word or one word phrase/collocation such as "What are your symptoms?"-"What are the side effects?" and “The defendant pleaded not guilty.”-“The defendant was found not guilty.”). Thus, they were tested on a list which consisted of the statements from both sources and new statements. Throughout the item test, the question “Have you seen this statement before?” appeared in blue on the upper left portion of the computer screen above the test sentences. Throughout the source test, however, the words “Who said:” appeared in red on the upper left portion of the computer screen above the test sentences, and both source pictures appeared side by side on the screen. Below each picture, the source label was printed in capital letters. The order of test sentence presentation alternated randomly for each participant. The experimental procedure is illustrated in Figure 1.

Participants decided on old/new recognition and source attribution sequentially; thus, they saw two options on the screen for item (i.e., YES and NO) and source query (i.e., JACK and SUSAN *or* JACK and JOHN), separately. They responded at their own pace by pressing the appropriate key. Crucially, as recommended for RT measurement in the additive-factor method, they were told to respond as quickly and as accurately as possible (cf. Shwartz et al., 1977; Van Duren & Sanders, 1988; Yap & Balota, 2007). Right-handed responses might be executed faster than left-handed responses by a right-handed participant (cf. Voss et al., 2010). Therefore, we determined certain keys which can be controlled only with one hand (for both left-handed and right-handed participants) and instructed participants to use their dominant hand. In the keyboard, “left arrow key” and “right arrow key” were assigned as “YES” and “NO”, respectively. If participants responded to the stimulus as old (i.e., “YES”), then, in order to indicate the source information, they were required to press “up arrow key” or “down arrow key” standing for “JACK” and “SUSAN” *or* “JOHN”, respectively. These answer choices and key assignments were shown again on the test screen. We additionally instructed participants how to place their fingers via a picture: The index finger was on the left arrow, and the ring finger was on the right arrow. They used their middle finger between these two keys to label the source (there was also a picture adjusted for the left-handed participants). We also told them to simply guess if they cannot remember whether the sentence was presented or not or who said the statement. Accuracy and response times (in ms) were automatically recorded via the lab.js scripts.

Results and Discussion

We fully followed our pre-registered plan for data exclusion and analysis. We acknowledge that RTs are noisy and that there is no agreed-upon outlier exclusion technique suitable for all experimental circumstances (also see Morís Fernández & Vadillo, 2020). Notably, Ulrich and Miller (1994) examined the effect of RT truncation on additivity of factor effects, and they found that truncation itself leads to even more confounded results than leaving spurious fast and slow answers in the analysis. Spurious answers cause noise in the data, but truncation may introduce a systematic bias. Moreover, calculating means from truncated data may turn out a really dangerous practice because RT distributions are positively skewed, resulting in exclusion of more RTs from the upper than the lower tail (Miller, 1991). Thus, given the lack of research on source monitoring latencies, and following the recommendations of Miller (1991) and Ulrich and Miller (1994), we did not specify a criterion for data trimming at the trial level. However, we draw conclusions based on the group means, and if there was a person deviating from the group mean considerably, this would affect the group mean and so the interpretation of the results. To avoid the influence of such extreme outliers, we excluded participants for all analyses (also for accuracy analyses) if their mean correct RTs for the item and/or source test is 2.5 or more standard deviations above or below the mean of their item retrieval difficulty × source retrieval difficulty condition (see Participants).

Accuracy

We applied the two-high threshold multinomial model of source monitoring (2HTSM; see Bayen et al., 1996) to separately estimate guessing-corrected item and source memory as well as several different response (guessing) biases (also see Bröder & Meiser, 2007). We aggregated response frequencies across participants for each item type (i.e., Jack [source A], Susan/John [source B], and new items) and then used the *multiTree* software (Moshagen, 2010) for parameter estimation, goodness-of-fit tests, and parameter comparisons. Parameter estimation was conducted through maximum-likelihood estimation methods, and we assessed model fit with the log-likelihood ratio goodness-of-fit statistic G^2. A nonsignificant G^2 indicates that the model fit the data. We evaluated parameter comparisons based on the chi-square difference test statistic ΔG^2. A significant ΔG^2 indicates that the parameter restrictions significantly worsened the model fit, and thus the corresponding parameter difference is considered significant. We set alpha to .05 for all tests.

Following Bayen et al. (1996), we analyzed the identifiable submodels of the 2HTSM by imposing certain parameter constraints. We first tested the most restrictive submodel 4, with four free parameters (i.e., $D_A = D_B = D_{New}$ [assuming equal item memory across sources and new distractor items], $d_A = d_B$ [assuming equal source memory across sources], b [item guessing], $a = g$ [assuming equal source guessing for recognized and unrecognized items]). However, this resulted in misfit in half of the conditions indicated by significant G^2 values (ps $< .001$). Next, we tested the less restrictive class of submodels 5, with five free parameters.

Submodels 5b and 5c assuming a difference in item memory across sources (either $D_A = D_{New} \neq D_B$ or $D_B = D_{New} \neq D_A$, respectively) also did not fit the data for half of the conditions (ps < .001). However, submodel 5a assuming equal source memory across sources ($d_A = d_B$) but a difference in source guessing for recognized and unrecognized items ($a \neq g$) and also submodel 5d assuming equal source guessing for recognized and unrecognized items ($a = g$) but a difference in source memory across sources ($d_A \neq d_B$) fit the data across the four conditions, all $G^2(1) \leq 2.57$, $p \geq .109$. For most conditions, submodels 5a and 5d resulted in the same model fit, but for the condition where participants generated items presented by the similar sources, fit was slightly better for 5a, $G^2(1) = 1.09$, $p = .297$, than 5d, $G^2(1) = 1.38$, $p = .241$. Our aim was to compare source memory across experimental conditions, not to compare source memory between sources (see also Riefer et al., 1994). In line with our research aim and the better fit of this submodel in one condition, we decided to use submodel 5a for our main analysis presented here. We additionally present parameter estimates of submodel 5d in Appendix, which gave consistent results with submodel 5a on the particulars of our research interest. Then, we performed joint model analyses to separately measure source-monitoring processes in each condition but in one overarching model allowing for comparison. The four-group joint submodel 5a fit the data across all four conditions well, $G^2(4) = 4.43$, $p = .351$.

The comparisons of greatest interest to us are on the parameter estimates for item and source memory (see Table 1; guessing parameters were also estimated but were not of central interest here). Firstly, we tested the effect of item retrieval difficulty on item memory by restricting parameter D to be equal across the generate and read groups of the same source retrieval difficulty levels. As expected, the model fit became significantly worse, $\Delta G^2(2) = 140.37$, $p < .001$. Follow-up analyses revealed that item memory was always better in the generate than the read condition, and this was significant both in the easy source retrieval condition, $\Delta G^2(1) = 36.92$, $p < .001$, and also in the difficult source retrieval condition, $\Delta G^2(1) = 103.45$, $p < .001$, implying a systematic effect of item encoding on item memory. Then, we tested the effect of item retrieval difficulty on parameter *d*, and these restrictions somewhat unexpectedly significantly decreased the model fit, $\Delta G^2(2) = 22.25$, $p < .001$. In the conditions where participants studied items from the similar sources, generating led to extremely, and unexpectedly, lower (actually around chance level) source memory than mere reading, $\Delta G^2(1) = 22.13$, $p < .001$. Consequently, albeit with its expected effect in the direction of better item memory for item generation, item encoding did not selectively affect item memory only.

Secondly, we tested the effect of source retrieval difficulty on source memory by restricting parameter *d* to be equal across the same item retrieval difficulty levels differing in source similarity. These restrictions significantly worsened the model fit, $\Delta G^2(2) = 144.56$, $p < .001$. Thus, we again followed up by implementing each restriction separately and observed that source memory was always better if participants studied items from the dissimilar sources, compared to the similar sources, both in the easy item retrieval condition, $\Delta G^2(1) = 130.97$, $p < .001$, and also in the difficult item retrieval condition, $\Delta G^2(1) = 13.59$, $p < .001$. Next, we tested the effect of source retrieval difficulty on parameter *D*, and these restrictions also significantly decreased the model fit, $\Delta G^2(2) = 21.67$, $p < .001$. On the conditions where participants studied items via mere reading, the similar sources (compared to the dissimilar ones) led to lower item memory, $\Delta G^2(1) = 21.41$, $p < .001$. Therefore, despite the expected results of better source memory in the dissimilar sources, source similarity also did not affect source memory selectively.

As pointed out earlier, selective effects are a crucial prerequisite for applying the additive-factor model. However, this applies to the item and source latencies (RTs), which are the dependent variable for the additive-factor analysis. Thus, next, we investigate whether the item and source memory manipulations selectively influenced item and source retrieval latencies.

Latency

As is the standard for the additive-factor method, we computed mean correct RTs (cf. Shwartz et al., 1977) for each participant in each of the four between-subjects conditions. Strictly speaking, using the raw RTs is the essence of the additive-factor method because the additive-factor logic applies to means of untransformed RTs only. Specifically, nonlinearly transformed RTs (e.g., log-RTs) or alternative summary measures (e.g., median RTs instead of mean RTs) would no longer behave additively even if different factors affect different serial stages selectively (Pachella, 1974; Sternberg, 1969, 1998).

We followed a set of typical analyses that are commonly performed in the additive-factor experiments. ANOVA is generally robust against violations of the normal distribution assumption (e.g., Maxwell & Delaney, 2004) and is the traditional way to assess additivity by testing main effects and interactions (cf. Shwartz et al., 1977; Townsend, 1984, 1990; Van Duren & Sanders, 1988; Yap & Balota, 2007). In the present study, our analyses were restricted to those trials in which participants gave both correct item and correct source responses. Thus, there was a clear focus on accuracy, and errors were excluded from the analysis (see Table 1, for error rates). Then, we performed separate 2 (item retrieval difficulty: easy [generating the study items], difficult [reading the study items]) × 2 (source retrieval difficulty: easy [dissimilar sources], difficult [similar sources]) between-subjects design ANOVAs for the item and source latency using JASP (Love et al., 2019). For the ANOVA analyses, we report partial eta squared (η_p^2) as the measure of effect size, and we rely on p values and set an alpha level of .05.

For the item RTs, neither the main effects of item retrieval difficulty, $F(1, 124) = 2.31$, $p = .131$, $\eta_p^2 = .02$, nor source retrieval difficulty, $F < 1$, nor their interaction, $F(1, 124) = 2.53$,

$p = .114$, $\eta_p^2 = .02$, was significant (see Figure 2A). Since item encoding had no definite effect on the item RTs, we could not meaningfully interpret the interaction test with regards to the question of serial or parallel processing.

For the source RTs, we observed a significant main effect of source retrieval difficulty, $F(1, 124) = 24.06$, $p < .001$, $\eta_p^2 = .16$, in the expected direction that the dissimilar sources (compared to the similar ones) led to faster source retrieval. The main effect of item retrieval difficulty, $F < 1$, and the item retrieval difficulty × source retrieval difficulty interaction, $F < 1$, were not significant (see Figure 2B). Thus, source processing had the desired definite and selective effect on source retrieval latencies. However, as noted before and preregistered, based on the source latency results, we cannot draw robust conclusions about the temporal course of item and source retrieval.

Discussion

In sum, findings revealed main effects of source retrieval difficulty on both source memory and source retrieval speed in the same direction. More specifically, we observed better source memory and faster source retrieval when the statements were presented by the dissimilar sources (Jack vs. Susan), compared to the similar sources (Jack vs. John). However, the main effect of item retrieval difficulty on item memory did not extend to item retrieval speed. In fact, generating led to better item memory than mere reading, as expected, but this effect did not extend to item retrieval speed, preventing subsequent additive-factor logic on the interaction test.

Notably, the latency processes underlying the source-monitoring paradigm have not been thoroughly investigated yet. In our first hypothesis, we deemed it theoretically plausible that effects of difficulty manipulations on memory performance should also extend to processing time. However, this did not fully hold in the current study. Although both manipulations affected item and source memory accuracy, respectively, this only translated into retrieval speed differences on the source test but not on the item test. Because a selective influence of item processing difficulty on item retrieval latency was not found, a crucial prerequisite for the additive-factor method was not met. We thus cannot test seriality versus parallelism of item and source retrieval with the current data. Nonetheless, this result provides interesting novel insights by showing a dissociation between effects on memory performance versus latency. Yet, the convergent evidence from source memory accuracy and source latency analyses suggests that the logical assumption underlying H1 may hold. Thus, to allow application of the additive-factor method on independent data, we conducted a second experiment with an improved design to increase our chances for selective effect on item retrieval latencies.

Experiment 2

In Experiment 2, we rethought and slightly modified our first experiment to better understand the previous item latency results. In the study phase of Experiment 1, the participants assigned to the easy item retrieval condition saw statements each containing one fragment, some letters of a word were missing (e.g., "Are you taking any other m_d_ _in_?"). However, the participants assigned to the difficult item retrieval condition saw statements each containing one underlined word and thus read them in the complete form (e.g., "Are you taking any other <u>medicine</u>?"). In the test phase, all participants, regardless of whether they generated or read the study items, were tested on targets and lures, which were also presented in the complete form (e.g., "Are you taking any other medicine?"). According to the encoding specificity principle, the encoding context affects what is stored, and hence, that affects to what extent retrieval cues are beneficial to have access to what is stored (see Tulving & Thomson, 1973). The lack of the desired difference between reading and generation on the item RTs thus might have been caused by slower access to the items in the generate condition due to the missing encoding-retrieval-context match. In terms of memory performance, the deep processing itself (here: generation) sustained retrieval against the context change (cf. Lockhart & Craik, 1990), still yielding better item memory, but the more sensitive RTs picked up the access difficulties due to the encoding-retrieval-context *mis*match. In contrast, the participants in the read condition saw the complete items both at study and test. Although the items were less deeply encoded, this reinstatement of the encoding context at test might have speed up item retrieval latencies.

As a remedy in Experiment 2, without changing the study phase, we presented all test items (both targets and lures) in fragmented form (e.g., "Are you taking any other m_d_ _in_?") for *all* conditions. Put differently, the participants in the generate condition saw the statements with the word fragments both at study and test, creating a match between encoding and retrieval. Thus, they should have an advantage in item retrieval—and retrieval speed—both from generating at encoding and the encoding-retrieval match at test. However, the participants in the read condition encoded the items by reading, but in the test phase, they had to complete the gaps (e.g., "medicine") first so that they could provide their item decisions. This mismatch between study and test might delay the item RTs for this group even further and serve to the desired difference between generation and reading in item latencies to be able to test the additive-factor logic. Thus, the read condition should have a disadvantage in item retrieval—and retrieval speed—both from shallow processing at encoding (here: reading) and the encoding-retrieval *mis*match at test.

Method

Crucially note that we maintained the same hypotheses, presented earlier, for this second study which was also pre-registered in the OSF. The preregistration protocol is available online at https://osf.io/pmye4/?view_only=bc343469b1ec45f7a0880ad632d3c679. All materials, experiment scripts, and results are available online at https://osf.io/9xhsq/?view_only=e72ee9004c914d0db0101565a568c699.

Participants

As we maintained the same hypotheses, the analysis plan followed the same rationale for sample size planning (i.e., 32 per item retrieval difficulty × source retrieval difficulty condition, so total $N = 128$). We tested 152 younger adults via Prolific. Monetary compensation and inclusion/exclusion criteria were the same as in Experiment 1. Data from fourteen participants were excluded due to their mean correct RTs for the item or source test being 2.5 or more standard deviations above or below than their item retrieval difficulty × source retrieval difficulty condition. Based on their feedback in the open text field or their inattentive performance on the encoding tasks (i.e., not answering half of the study trials) or else due to technical problems, we also omitted data from six participants who did not conduct our study carefully. We excluded one participant who did not show item memory performance above-chance level (i.e., hit rates > false alarm rates). We also excluded three participants whose source memory performance (i.e., ACSIM) was 2.5 or more standard deviations below the mean performance of their assigned condition. Analyses were carried out with the remaining 128 (64 female, 64 male; M_{age}= 25.31 years, age range = 18 – 30).

Design, Materials, and Procedure

Materials (agents and statements serving as sources and items, respectively) and design were the same as in Experiment 1. We also kept the procedure of the first experiment except for the following crucial change. As mentioned above, in the test phase, we presented all test items (both targets and lures) in fragmented form for all conditions, independent of whether participants encoded the study items with generation versus mere reading or by the dissimilar versus similar sources (see Figure 1). For the fragmented *targets*, we used the word fragments from the study phase of Experiment 1 (previously used for the generate condition; e.g., “Are you taking any other m_d_ _in_?”). For the fragmented *lures*, a new set of word fragments[2] was arranged corresponding to the distracting versions of the studied target statements (e.g., “Are you taking any other pr_ _cr_p_ _ _ _s?”).

Results and Discussion

Accuracy

As in Experiment 1, we used the 2HTSM (Bayen et al., 1996) for the statistical analyses of guessing-corrected memory accuracy. The test settings remained the same. We first implemented the most restrictive submodel 4, with four free parameters (i.e., $D_A = D_B = D_{New}$ [assuming equal item memory across sources and new distractor items], $d_A = d_B$ [assuming equal source memory across sources], b [item guessing], $a = g$ [assuming equal source guessing for recognized and unrecognized items]), resulting in misfit in half of the conditions indicated by the significant G^2 values (ps $\leq .005$). Next, we tested the less restrictive submodel 5 variants, with five free parameters. Unlike Experiment 1, submodels 5a, which assumes equal source memory across sources ($d_A = d_B$) but a difference in source guessing for recognized and unrecognized items ($a \neq g$), and 5d, which assumes equal source guessing for recognized and unrecognized items ($a = g$) but a difference in source memory across sources ($d_A \neq d_B$), did not fit the data for half of the conditions (ps $\leq .007$). In contrast, both 5-parameter submodels restricting source memory ($d_A = d_B$) and source guessing ($a = g$) to be equal fit the data from all four conditions well: Submodel 5b (additionally restricting $D_A = D_{New} \neq D_B$), all $G^2(1) \leq 3.01$, $p \geq .083$, and submodel 5c (additionally restricting $D_B = D_{New} \neq D_A$), all $G^2(1) \leq 3.12$, $p \geq .077$. Regarding their differential restrictions on the item memory parameters (see above), Bell et al. (2015) argued that both submodels violate the assumption that the probability of recognizing distractors as “new” mirrors the probability of recognizing targets as “old” (i.e., $D_{Old} = D_{New}$). Thus, when item memory differs between sources ($D_A \neq D_B$), as is the case for the current data, Bell et al. suggest setting $D_{New} = (D_A + D_B) / 2$, as a remedy, and they previously showed the successful application of this restriction across different experiments (also see Bell et al., 2010, 2013). Thus, we implemented this item memory restriction alongside the source memory and source guessing restrictions made by both of these submodels ($d_A = d_B$; $a = g$), which fit the data for the four conditions well, all $G^2(1) \leq 3.07$, $p \geq .080$. Surprisingly, in the read conditions (regardless of whether the sources were dissimilar or similar), item memory was higher for the items said by Jack than the items said by Susan/John, both $\Delta G^2(1) \geq 7.87$, $p \leq .005$. We refrain from further interpreting this unexpected difference in item memory by sources as it has no implications for the subsequent additive-factors RT analysis of primary interest. Also note that this item memory difference between sources did not occur in Experiment 1 where we used the same sources and item statements.

Next, we conducted the joint model analyses to compare the separately estimated parameters from the different conditions in an overarching model, and the four-group joint model fit the data well, $G^2(4) = 5.80$, $p = .215$. Again, critical to our research question, we compared the parameter estimates for item and source memory (see Table 2, also for the guessing parameters). Interested readers can find all analyses on response biases as well as the submodel tests in the OSF. Firstly, we tested the effect of item retrieval difficulty on item memory by restricting parameters D_A and D_B to be equal across the generate and read groups of the same source retrieval difficulty levels, and as to be expected, the model fit became significantly worse, $\Delta G^2(4) = 413.79$, $p < .001$. Follow-up analyses replicated Experiment 1 and revealed that item memory was always better in the generate than the read condition, and this was significant in all comparisons, all $\Delta G^2(1) \geq 56.18$, $p < .001$. However, unlike Experiment 1 but in line with the expected and desired selective influence manipulation, when we tested the effect of item retrieval difficulty on parameter d, we found no significant difference for source memory, $\Delta G^2(2) = 0.17$, $p = .917$.

Secondly, we tested the effect of source retrieval difficulty on source memory by restricting parameter *d* to be equal across the same item retrieval difficulty levels differing in source similarity. These restrictions significantly worsened the model fit, $\Delta G^2(2) = 109.80$, $p < .001$. Thus, we again followed up by implementing each restriction separately. Similar to Experiment 1, we observed that source memory was always better if participants studied items from the dissimilar sources, compared to the similar sources, both in the easy item retrieval condition, $\Delta G^2(1) = 83.36$, $p < .001$, and also in the difficult item retrieval condition, $\Delta G^2(1) = 26.45$, $p < .001$. However, unlike Experiment 1 but again in line with the expected and desired selective influence manipulation, when we tested the effect of source retrieval difficulty on parameters D_A and D_B, we found no significant difference for item memory, $\Delta G^2(4) = 1.04$, $p = .904$.

Findings indicated that item encoding had the desirable selective effect on item memory, leaving source memory unaffected. Also as intended, source similarity had a consistent effect on source memory but no significant effect on item memory. Furthermore, the floor level source memory previously observed in one of the conditions from Experiment 1 where participants generated items presented by similar sources disappeared. Therefore, in Experiment 2, we had above-chance source memory for all conditions although, in line with our difficulty manipulation, the similar sources always led to lower source memory. In the following section, we investigate whether the selective item and source memory accuracy influences translate into selective item and source retrieval speed influences, before assessing our preregistered hypotheses on the interaction effect.

Latency

The analysis plan and test settings were identical to Experiment 1. Again, we restricted our analyses to only correctly solved trials (see Table 2, for error rates) and then conducted separate 2 (item retrieval difficulty: easy [generating the study items], difficult [reading the study items]) × 2 (source retrieval difficulty: easy [dissimilar sources], difficult [similar sources]) between-subjects design ANOVAs for the item and source latency, respectively.

For the item RTs, unlike Experiment 1, we observed the significant main effect of item retrieval difficulty, $F(1, 124) = 11.16$, $p = .001$, $\eta_p^2 = .08$, such that generating (compared to reading) led to faster item retrieval. The main effect of source retrieval difficulty, $F < 1$, and the item retrieval difficulty × source retrieval difficulty interaction , $F < 1$, were not significant (see Figure 3A). Given that item encoding had a definite and selective effect on the item RTs, a nonsignificant interaction can be interpreted according to the additive-factor logic as supporting serial processing stages of item and source retrieval.

For the source RTs, we, again, observed the significant main effect of source retrieval difficulty, $F(1, 124) = 13.63$, $p < .001$, $\eta_p^2 = .10$, such that the dissimilar sources (compared to the similar ones) led to faster source retrieval. The main effect of item retrieval difficulty, $F < 1$, and the item retrieval difficulty × source retrieval difficulty interaction , $F < 1$, were not significant (see Figure 3B). Critically, the source latency results of Experiment 1 were replicated in Experiment 2.

Discussion

Experiment 2 successfully replicated the selective influences of item generation on item (but not source) memory (Mulligan et al., 2006) and source similarity on source (but not item) memory (Bayen et al., 1996). These clear effects on memory accuracy provided ideal conditions for also observing effects on latencies, including the item latencies which through the improved design of Experiment 2 should no further be hampered by the lack of encoding-retrieval-context match. Indeed, as predicted, reinstating the original *item* encoding context at test did not alter the pattern in the source latencies in comparison to Experiment 1 but specifically in the item latencies. More specifically, this remedy elicited the desired difference between item reading and item generation on the item RTs (reading RTs > generation RTs). It also seems to have promoted the selective effects of item encoding and source similarity on item and source memory performance, respectively, whereas in Experiment 1 these effects emerged but were not selective.

As in Experiment 1, Experiment 2 again showed the consistent and selective effect of source retrieval difficulty on both source memory and source retrieval speed, suggesting that dissimilar sources do not only improve source memory but render faster source retrieval relative to similar sources. Although, as reasoned in the introduction, such an effect is not very informative for the question on seriality versus parallelism, it yet at the same time provides novel empirical evidence about the association between source memory accuracy and the latency of source retrieval conditional on item retrieval.

Crucial to our hypotheses, the definite and selective effects of the item and source manipulations on their respective RTs allowed for interpreting the interaction effect with regard to seriality versus parallelism of item and source retrieval according to the additive-factors logic. According to Sternberg (2013), "the prediction of additivity should be thought of as depending on a two-part hypothesis: stages (seriality) and selective influence. Observation of additivity supports both parts, just as confirmation of a prediction from any theory supports that theory." (p. 2). Thus, the additivity of item and source effects on item latencies observed here, with no evidence for their interaction, favors seriality. Combined with the significant source similarity main effect on source RTs but not item RTs, this indicates that source retrieval starts after the item response has been given.

General Discussion

The primary focus of this paper was to investigate whether retrieval processes for an item and its source operate in strict sequence or in parallel in the widely used standard sequential source-monitoring test (i.e., first probing whether an item is old or new and for items judged old consecutively probing for the source). We applied the additive-factor method herein to test the alternative mental architectures (serial vs. parallel processing) in source monitoring based on raw RT data. Accordingly, additive factors effects on mean RT (i.e., significant main effects of both factors but no interaction) support a serial arrangement of stages together with selective influences of factors on the corresponding stages (Sternberg, 1998, 2013). Adopting this logic to the current experimental paradigm, one needs to identify factors having definite and selective effects on the duration of item and source processing. Several behavioral findings have shown the dissociability of item and source memory accuracy with selective effects (e.g., Lindsay & Johnson, 1991; Raye & Johnson, 1980), but what (selectively) affects the *speed* of item and source retrieval is currently not well understood. On the basis of the selective effects observed on item (Mulligan et al., 2006) and source memory (Bayen et al., 1996) accuracy, we manipulated item encoding and source similarity and ultimately expected the transfer of their effects on memory accuracy to latency. In Experiment 1, the dissimilar sources (compared to the similar sources) resulted in faster source retrieval in addition to the performance benefit, but the expected performance benefit of item generation (compared to item reading) did not render faster RTs, hindering further test of additivity. In Experiment 2, we modified the test phase in line with the encoding specificity principle and observed comparable accuracy results, but this time leading to cleaner interpretations of selective influences on item and source memory. These selective influences further analogously transferred to the RTs of the specific component as a main effect, allowing a more precise interpretation of the interaction test on the item RTs. Non-interacting effects of item encoding with source similarity on the item latency suggest that source similarity does not affect the duration of item encoding stage. Thus, we demonstrated that item and source processing proceed in serial succession in the standard sequential source-monitoring paradigm, consistent with the order they were probed. Even though participants were tested next for sources upon their "old" answer, they did not start the source retrieval during the item test, and only item retrieval took place at this stage. The source retrieval was rather started consecutively during the source test.

Earlier studies provided evidence in favor of slower recollection processes compared to more automatic familiarity processes (e.g., Jacoby, 1991; McElree et al., 1999). Critically, the present experiments found that the source RTs were always faster than the item RTs regardless of the conditions, although source identification tasks require more recollection demands (Yonelinas, 1999). Considering this systematic difference between the item and source RTs, one may conclude that source retrieval is faster than item retrieval, which would be inconsistent with earlier response-signal studies. It is worth noting that the actual total time devoted to the source attributions was separated here with our sequential test design, and Johnson et al. (1994) already showed that the availability of source information takes longer than item information in total. However, most importantly, source retrieval may nonetheless be started in parallel to item retrieval or serially. Here, the key benefit of the additive-factor method is that, instead of comparing the stages to each other, it manipulates the stages selectively with a sophisticated factorial design and thus provides more fine-grained evidence about their mental organizations, which favored seriality for the current interest.

The next question now arises as to why the primarily recollection-based source RTs were faster in the current study. An important point worthy of further attention could be the standard sequential source-monitoring paradigm itself and, more specifically, the interplay between item and source decisions therein. We acknowledge that the current source RTs were confined to the conditional source judgments as standard in source-monitoring research. Consecutive source testing after the state of item recognition might have eased source retrieval (see Tanyas & Kuhlmann, 2023, for a similar argument). This prediction also fits the general assumption of multinomial processing tree (MPT) models for source monitoring (for an extensive review, see Erdfelder et al., 2009; for a tutorial, see Schmidt et al., 2023) that source discrimination is only possible in the state of item recognition. However, in a novel test design, for example, if the source identification task is given separately from, or without any preordered, item detection process (cf. Osth et al., 2018; Starns et al., 2008), this difference in item and source RTs might vanish or even reverse. Indeed, when investigating the mouse movements to response options in both tests, we found more straightforward source trajectories in the standard sequential test as also implemented in the current design (with faster source RTs) but the source trajectories were more curved than item trajectories, indicating greater difficulty of the source retrieval, when item and source tests were blocked further apart in time. However, such an imposed time delay between the item and source test forces seriality by design and thus is not informative regarding the possibility of parallel item and source retrieval. Although less severe, one might also object that the standard sequential test of item immediately followed by source somewhat imposes a seriality. Alternatively, source monitoring can be tested in one step, asking participants to judge if an item was presented by Source A, Source B, or is new on one screen. In this case, a source attribution implies that the item is perceived to be old but this is not assessed separately. As a consequence, only one latency can be derived from this test, preventing the application of the additive-factor logic. Thus, we fully acknowledge that even if individuals do not engage in parallel retrieval of item and source in specific circumstances such as the sequential test employed here, that does not necessarily mean they could not do so at all. Furthermore, under some circumstances, even above-chance source memory for unrecognized items is found (see Fox & Osth, 2022, for an overview), and it would also be interesting to investigate in these paradigms to what extent item information is also already retrieved in parallel (but not yet finished) when the source decision is rendered. Thus, future research should investigate boundary conditions to the seriality of item and source retrieval observed here with a challenge for this research being how to assess item and source decisions closely in time yet with separable latencies.

It is important to consider that the raw RTs relied upon here in line with the standard approach for the additive-factor method (cf. Sternberg, 1998) of course do not purely reflect the latencies of item and source decision processing but also non-decision processes of stimulus perception and psychomotor response activation (cf. Voss et al., 2010; see also Tanyas et al., 2023). Critically, however, this non-decision component was unaffected by the implemented item encoding and source similarity manipulation because the stimulus display during test and the response keys remained constant across conditions. However, another possible explanation of the faster source RTs (relative to the item RTs) could be a general reduction in this non-decisional time during the source test, and that might have emerged as an artifact of consecutive responding. In fact, there was no interval between the item and source test, and participants knew that the same word would be tested again next upon their "old" answer. Thus, perceiving the same stimulus again likely did not take as much time in the source as in the preceding item test. Future research is needed to disentangle decisional and non-decisional processes underlying the noisy RTs of such a complex memory task.

Importantly, even though our manipulations affected item and source memory at the accuracy of responding, we are not necessarily claiming that this is a single process (i.e., memory) behind item and source RTs, but rather, can also be several processes, including decision and guessing processes. Indeed, based on the source-monitoring framework and the employed multinomial model which relies thereon, we must assume that memory and guessing processes contribute to the item and source responses in the source-monitoring task. As long as they can be definitely tied to item and source processing and are thus not affected by the other manipulation, that is, the selective influence is given, as was the case in Experiment 2, we can use the additive-factor logic herein to gain more insight into item and source retrieval. Additionally, our approach involved assessing only targets and excluding participants with particularly low memory performance in order to strive that memory contributes to the latencies, thereby ensuring that analyzed RTs at least tap into memory speed.

One could object that our source retrieval difficulty manipulation was not strong enough to produce the effect on the item latencies, but that can be rejected by pointing to the consistent source latency results in both experiments. Obviously, source similarity induces an effect on the source RTs, but it did not appear in the item RTs. We give prominence to exactly this point for the current research interest. With no additional source effects on the item latencies, we found no convincing evidence for the fact that in the item test people already begin to process source information at some point. That rather suggests strictly serial processing stages, and the item response is the boundary between the end of item processing and the beginning of source processing.

Apart from the temporal ordering of item and source retrieval, the current experiments have raised the question of whether better memory renders faster retrieval. In both experiments, we obtained evidence in the source test with better source memory and faster source retrieval for dissimilar sources compared to similar sources. In the item test, however, Experiment 1 showed the dissociation between effects on memory performance versus latency while Experiment 2 showed the convergent effect of item generation both on item memory and item retrieval speed after controlling for the encoding-retrieval-context match. Importantly, we are not claiming that item processing is more fragile than source processing. Instead, the encoding-retrieval mismatch in Experiment 1 resulted from the manipulation of item encoding and thus did not selectively influence item judgment latencies. Carefully note that across experiments, the source context, that is, the picture of the face accompanied by the source label, was presented both at encoding and at test. Therefore, the consistent source effects in Experiments 1 and 2 might be the result of the preserved source context both at encoding and retrieval. Indeed, we suspect that source latencies would be sensitive to altering aspects of the source presentation at test (cf. Symeonidou & Kuhlmann, 2021). Thus, similar dissociations between source accuracy and source latency seem possible in case of encoding-retrieval mismatch in the source test.

If encoding and retrieval context conditions match, better item memory appears to be associated with faster correct item RTs and better source memory with faster correct source RTs, as shown in Experiment 2. What does this mean with respect to the processing assumptions underlying the 2HTSM model (Bayen et al., 1996) which we used to measure the accuracies of item memory (i.e., *D*) and source memory (i.e., *d*)? According to this model, correct source attributions to target items that entered in our additive-factor analysis basically emerge in two different ways: (1) via memory processes and (2) via lucky guessing. Taking this into account, faster observed RTs do not necessarily imply that the speed of one or both latent processing routes has increased. It is possible that (1) memory-based responding is fast and (2) judgment involving guessing is generally slow, while the speed of both processing routes is unaffected by our experimental manipulations. Still, we would expect faster overall correct RTs when memory is higher, simply because a larger proportion of the overall RT distribution is determined by fast memory-based responses (cf. Heck & Erdfelder, 2016, 2020). Most important in our present context, this does not invalidate application of the additive-factor logic in any way. Applying the same logic, if source memory would be involved already before the item response is given, then better source memory should increase the proportion of fast responses among the total item RT distribution and thus show influences of the source manipulation on item RT. The absence of any effects of the source manipulation on item RTs supports that source retrieval comes into play only after the item response has been provided. Future research should attempt a more thorough and detailed investigation of the latency mechanisms underlying item and source responses to examine whether the employed accuracy manipulations alter the speed of memory.

A challenging limitation of research on memory latencies in source monitoring is that source monitoring is a complex memory task and thus produces more incorrect responses than the standard short-term memory-scanning task that Sternberg (1969) originally applied the additive-factor logic to. More relevant to our primary hypotheses, however, the observed error rates (i.e., the proportions of missed targets and source attribution errors) increased in parallel to our difficulty manipulations, providing additional support that the manipulations worked as intended. This notwithstanding, we still found no evidence for an interaction of these clearly impactful factors on mean RT.

To our best knowledge, the current work is the first application of the additive-factor method to source monitoring. With its emphasis on factorial interactions and selective influences, the additive-factor method certainly merits attention, but additional tests of seriality (vs. parallelism) based on a *joint* consideration of RTs and accuracy (cf. Townsend, 1990) would be desirable in future source-monitoring studies.

Declarations

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant KU3329/1-1, project number 392884168. Hilal Tanyas was further supported by the University of Mannheim Graduate School of Economic and Social Sciences (GESS).

Conflicts of interest

The authors report no conflict of interest.

Ethics approval

All procedures performed in the study were in accordance with the principles of the Declaration of Helsinki, the guidelines of the German Psychological Society (DGPs), and the guidelines of the University of Mannheim ethics committee.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Availability of data and materials

The datasets analyzed during the current study are available in the OSF repository and can be accessed via the at https://osf.io/9xhsq/?view_only=e72ee9004c914d0db0101565a568c699.

Code availability

The analysis codes and experiment scripts can be accessed via the OSF at https://osf.io/9xhsq/?view_only=e72ee9004c914d0db0101565a568c699.

References

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of Sternberg's method. *Psychological Review*, *123*(5), 481-509. https://doi.org/10.1037/rev0000030

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., & Kuhlmann, B. G. (2011). Influences of source–item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1-17. https://doi.org/10.1016/j.jml.2010.09.001

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

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*(3), 480–500. https://doi.org/10.3758/BF03198562

Bell, R., Buchner, A., & Musch, J. (2010). Enhanced old–new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition*, *117*(3), 261-275. https://doi.org/10.1016/j.cognition.2010.08.020

Bell, R., Giang, T., Mund, I., & Buchner, A. (2013). Memory for reputational trait information: Is social–emotional information processing less flexible in old age? *Psychology and Aging*, *28*(4), 984–995. https://doi.org/10.1037/a0034266

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 456–472. https://doi.org/10.1037/xlm0000034

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old–new recognition. *Emotion*, *17*(1), 120–130. https://doi.org/10.1037/emo0000211

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*(1), 52-60. https://doi.org/10.1027/0044-3409.215.1.52

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

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

Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review, 23*(5), 1440–1465. https://doi.org/10.3758/s13423-016-1025-6

Heck, D. W., & Erdfelder, E. (2020). Benefits of response time-extended multinomial processing tree models: A reply to Starns (2018). *Psychonomic Bulletin & Review, 27*(3), 571–580. https://doi.org/10.3758/s13423-019-01663-0

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

Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language*, *48*(3), 489-501. https://doi.org/10.1016/S0749-596X(02)00537-5

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541. https://doi.org/10.1016/0749-596X(91)90025-F

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

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

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* [Unpublished doctoral dissertation]. New York University.

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1365-1376. https://doi.org/10.1037/a0028147

Lindsay, D. S. (2008). Source monitoring. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 325-348). Elsevier. https://doi.org/10.1016/B978-012370509-9.00175-3

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

Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology / Revue canadienne de psychologie*, *44*(1), 87–112. https://doi.org/10.1037/h0084237

Lopes, E. J., & Garcia, R. B. (2014). On the possibility of using reaction time to study false memories. *Psychology & Neuroscience*, *7*(3), 393-397. https://doi.org/10.3922/j.psns.2014.047

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(2), 1–17. https://doi.org/10.18637/jss.v088.i02

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

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). The effect of context variability on source memory. *Memory & Cognition*, *34*(8), 1578-1586. https://doi.org/10.3758/BF03195921

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*. Lawrence Erlbaum Associates.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287-330. https://doi.org/10.1037/0033-295X.86.4.287

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

Miller, J. (1991). Short report: Reaction time analysis with outlier exclusion: Bias varies with sample size. *The Quarterly Journal of Experimental Psychology*, *43*(4), 907-912. https://doi.org/10.1080/14640749108400962

Miller, J. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review*, *100*(4), 702-715. https://doi.org/10.1037/0033-295X.100.4.702

Morís Fernández, L., & Vadillo, M. A. (2020). Flexibility in reaction time analysis: Many roads to a false positive? *Royal Society Open Science*, *7*, 190831. http://dx.doi.org/10.1098/rsos.190831

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42-54. https://doi.org/10.3758/BRM.42.1.42

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 838-855. https://doi.org/10.1037/0278-7393.30.4.838

Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 836-846. https://doi.org/10.1037/0278-7393.32.4.836

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). 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

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Lawrence Erlbaum Associates.

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

Raye, C. L., & Johnson, M. K. (1980). Reality monitoring vs. discriminating between external sources of memories. *Bulletin of the Psychonomic Society*, *15*(6), 405-408. https://doi.org/10.3758/BF03334572

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

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 680–693. https://doi.org/10.1037/0278-7393.20.3.680

Schmidt, O., Erdfelder, E., & Heck, D. W. (2023). How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000561

Schweickert, R., Fisher, D. L., & Goldstein, W. M. (2010). Additive factors and stages of mental processes in task networks. *Journal of Mathematical Psychology*, *54*(5), 405-414. https://doi.org/10.1016/j.jmp.2010.06.004

Schweickert, R., & Townsend, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology*, *33*(3), 328-347. https://doi.org/10.1016/0022-2496(89)90013-8

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

Shwartz, S. P., Pomerantz, J. R., & Egeth, H. E. (1977). State and process limitations in information processing: An additive factors analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(3), 402-410. https://doi.org/10.1037/0096-1523.3.3.402

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

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

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

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (pp. 703-863). MIT Press.

Sternberg, S. (2013). The meaning of additive reaction-time effects: Some misconceptions. *Frontiers in Psychology*, *4*, 744. https://doi.org/10.3389/fpsyg.2013.00744

Symeonidou, N. & Kuhlmann, B. G. (2021) A novel paradigm to assess storage of sources in memory: The source recognition test with reinstatement, *Memory*, *29*(4), 507-523. https://doi.org/10.1080/09658211.2021.1910310

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

Tanyas, H., Liss, J. V., & Kuhlmann, B. G. (2023). *Information accumulation on the item versus source test of source monitoring: Insights from diffusion modeling* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, *28*(4), 363-400. https://doi.org/10.1016/0022-2496(84)90007-5

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*(1), 46-54. https://doi.org/10.1111/j.1467-9280.1990.tb00067.x

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*(1), 34-80. https://doi.org/10.1037/0096-3445.123.1.34

Van Duren, L. L., & Sanders, A. F. (1988). On the robustness of the additive factors stage structure in blocked and mixed choice reaction designs. *Acta Psychologica*, *69*(1), 83-94. https://doi.org/10.1016/0001-6918(88)90031-5

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

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 274-296. https://doi.org/10.1037/0278-7393.33.2.274

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

Footnotes

[1] We did not use the original doctor-lawyer paradigm because that paradigm was employed to understand the influence of prior knowledge on source judgements in source monitoring. Instead, we tailored the doctor-lawyer paradigm to our research question for which it is well-suited because it provides similar distractor statements focusing on the difference of one word/short phrase that could be used in the generation manipulation. Further, this paradigm does not require audio, making it more accessible as an online study. We therefore changed the number and the arrangements of the statements (see Materials). We did not use the equally expected statements from this paradigm because they were too few but have been found to be remembered better than both expected doctor and expected lawyer statements (Bayen et al., 2000; Bayen & Kuhlmann, 2011) so they could not be mixed.

[2] We manually arranged these fragments (not randomly) by considering that they should be neither too difficult nor too easy. If all word fragments were too difficult, participants might have been frustrated, which is undesirable in terms of motivational reasons. On the contrary, if all word fragments were too easy, that might have caused little or no difference between the read and generate condition, and this would not elicit the intended effect on item memory. Therefore, we assessed the optimal difficulty of the word fragments used in both Experiments 1 and 2 with prior pilot testing.

Table 1

Error Rates and Parameter Estimates of the 2HTSM for Experiment 1

		Error Rate	**Memory Parameters**		Guessing Parameters			Model Fit
Item Encoding	Source Similarity	(out of targets)	***D***	***d***	*b*	*a*	*g*	$G^2(1)$
generate	dissimilar	0.39 (0.13)	**.63 [.60, .66]**	**.65 [.58, .72]**	.51 [.47, .55]	.72 [.60, .84]	.39 [.33, .45]	2.57
	similar	0.60 (0.10)	**.62 [.59, .65]**	**.02 [.00, .10]**	.47 [.43, .51]	.49 [.45, .54]	.53 [.47, .60]	1.09
read	dissimilar	0.50 (0.16)	**.49 [.46, .53]**	**.63 [.54, .72]**	.41 [.37, .44]	.84 [.69, 1.00]	.37 [.31, .43]	.01
	similar	0.62 (0.13)	**.38 [.34, .41]**	**.35 [.24, .47]**	.41 [.38, .44]	.60 [.50, .71]	.49 [.43, .54]	.75

Note. Standard deviations for error rates are in parentheses. The presented table includes parameter estimates of submodel 5a of the two-high threshold multinomial model of source monitoring (2HTSM; Bayen et al., 1996). With 1 *df*, a good model fit (i.e., $p > .05$) is given for all conditions. *D* = probability of detecting a specific item as old (or a distractor as new); *d* = probability of correctly discriminating the source of an item; *b* = probability of guessing that an item is old; *a* = probability of guessing that a detected item was presented by Jack; *g* = probability of guessing that an undetected item was presented by Jack. Brackets indicate 95% confidence intervals that were truncated at the parameter boundaries of 0 and 1.

Table 2

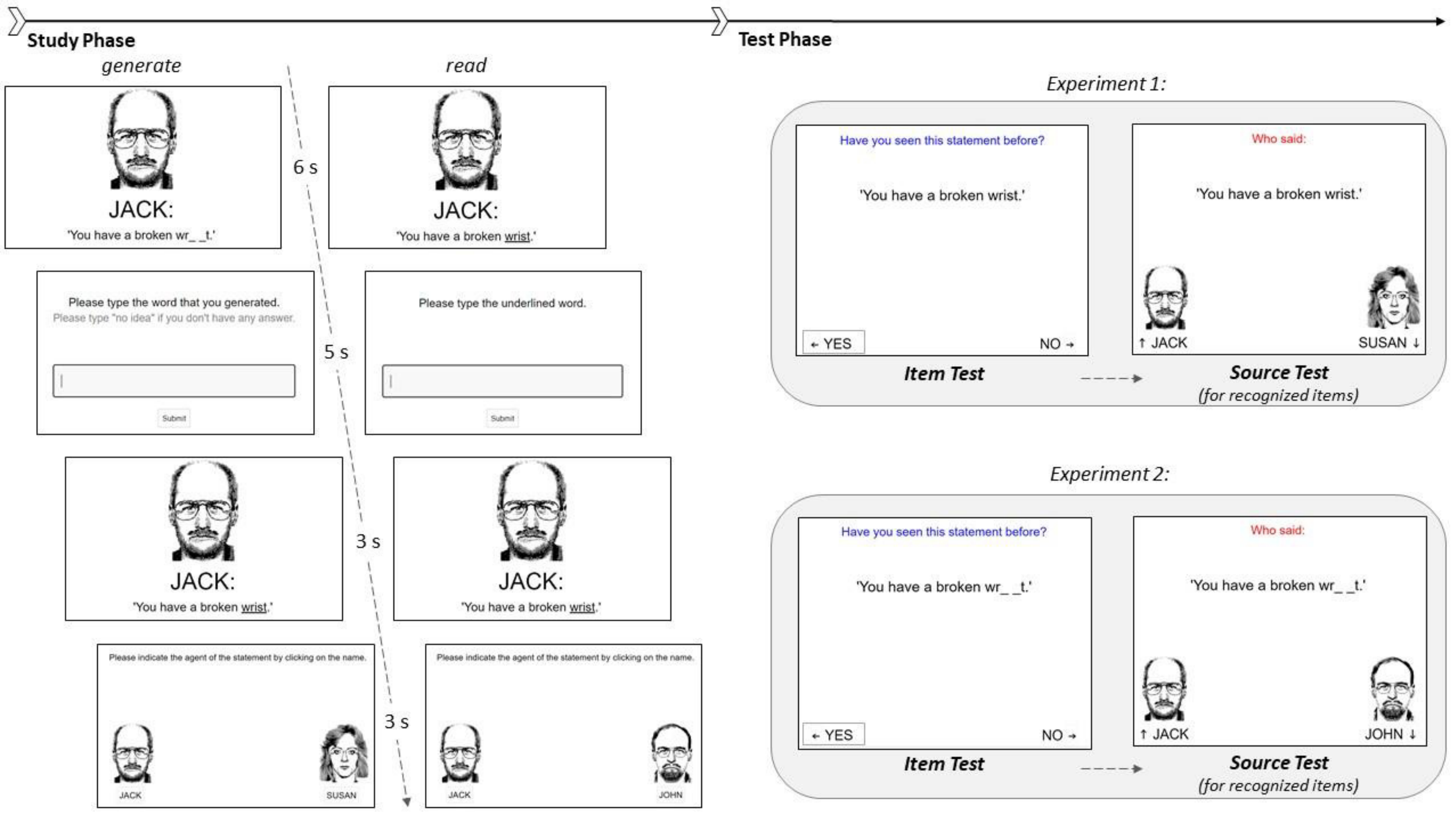
Error Rates and Parameter Estimates of the 2HTSM for Experiment 2

		Error Rate	**Memory Parameters**			Guessing Parameters		Model Fit
Item Encoding	Source Similarity	(out of targets)	D_{Jack}	$D_{Susan/John}$	d	b	g	$G^2(1)$
generate	dissimilar	0.34 (0.12)	**.73 [.69, .78]**	**.70 [.65, .74]**	**.64 [.58, .70]**	.49 [.45, .54]	.52 [.48, .57]	3.07
	similar	0.50 (0.11)	**.74 [.69, .78]**	**.69 [.65, .74]**	**.20 [.13, .27]**	.48 [.43, .53]	.50 [.47, .53]	< .01
read	dissimilar	0.56 (0.12)	**.46 [.41, .52]**	**.29 [.24, .35]**	**.64 [.52, .75]**	.41 [.38, .44]	.48 [.44, .52]	.01
	similar	0.66 (0.12)	**.45 [.40, .50]**	**.33 [.28, .39]**	**.23 [.12, .33]**	.34 [.31, .37]	.50 [.46, .53]	2.71

Note. Standard deviations for error rates are in parentheses. The presented model implements restrictions from submodels 5b and 5c of the two-high threshold multinomial model of source monitoring (2HTSM; Bayen et al., 1996), but the probability of detecting distractors is set to $D_{New} = (D_{Jack} + D_{Susan/John}) / 2$ (Bell et al., 2015). With 1 *df*, a good model fit (i.e., $p > .05$) is given for all conditions. D_{Jack} & $D_{Susan/John}$ = probability of detecting an item presented by Jack or Susan/John, respectively; d = probability of correctly remembering the source of an item; b = probability of guessing that an item is old; g = probability of guessing that an item was presented by Jack. The model parameters are probability estimates that can range from 0 to 1. Brackets indicate 95% confidence intervals.

Figure 1

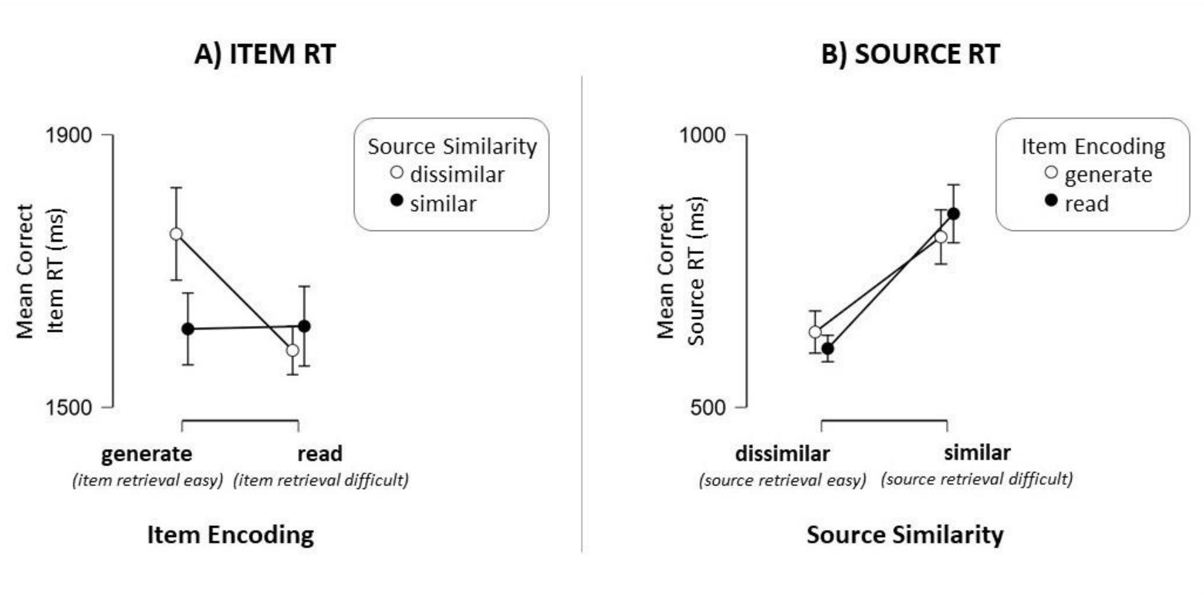
Procedure for Experiments 1 and 2



Note. Experiments 1 and 2 involve 4 combinations of item (generate vs. read) and source retrieval difficulty (Jack and Susan vs. Jack and John). For illustrative purposes, here we only present two combinations of them (generate with dissimilar sources and read with similar sources) but in the study design they were fully crossed.

Figure 2

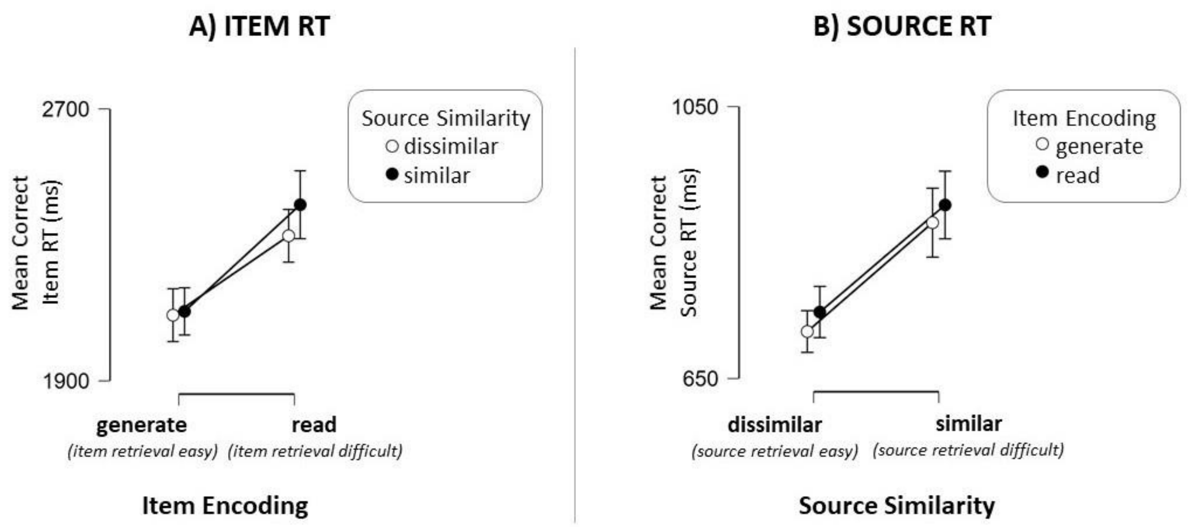
Mean Correct Item and Source RTs for Experiment 1



Note. Error bars represent the standard error of the mean.

Figure 3

Mean Correct Item and Source RTs for Experiment 2



Note. Error bars represent the standard error of the mean.

Appendix

Table A1

Parameter Estimates of the 2HTSM for Experiment 1 With Submodel 5d

		Memory Parameters			Guessing Parameters		Model Fit
Item Encoding	Source Similarity	$\boldsymbol{D}$	$\boldsymbol{d_{Jack}}$	$\boldsymbol{d_{Susan/John}}$	b	g	$G^2(1)$
generate	dissimilar	**.63 [.60, .66]**	**.84 [.75, .93]**	**.35 [.15, .55]**	.51 [.47, .55]	.39 [.33, .45]	2.57
	similar	**.62 [.59, .65]**	**.00 [.00, .20]**	**.05 [.00, .24]**	.47 [.43, .51]	.52 [.45, .58]	1.38
read	dissimilar	**.49 [.46, .53]**	**.91 [.80, 1.00]**	**.15 [.00, .41]**	.41 [.37, .44]	.37 [.31, .43]	.01
	similar	**.38 [.34, .41]**	**.50 [.30, .70]**	**.20 [.00, .43]**	.41 [.38, .44]	.49 [.43, .54]	.75

Note. The presented table includes parameter estimates of submodel 5d of the two-high threshold multinomial model of source monitoring (2HTSM; Bayen et al., 1996). With 1 *df*, a good model fit (i.e., $p > .05$) is given for all conditions. D = probability of detecting a specific item as old (or a distractor as new); d_{Jack} & $d_{Susan/John}$ = probability of correctly remembering the source of an item presented by Jack or Susan/John, respectively; b = probability of guessing that an item is old; g = probability of guessing that an item was presented by Jack. Brackets indicate 95% confidence intervals that were truncated at the parameter boundaries of 0 and 1.

Information Accumulation on the Item Versus Source Test of Source Monitoring: Insights From Diffusion Modeling

Hilal Tanyas, Julia V. Liss, and Beatrice G. Kuhlmann

Department of Psychology, University of Mannheim

Author Note

Hilal Tanyas https://orcid.org/0000-0002-5849-677X

Julia V. Liss https://orcid.org/0009-0008-4914-4175

Beatrice G. Kuhlmann https://orcid.org/0000-0002-3235-5717

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant KU3329/1-1, project number 392884168. Hilal Tanyas was further supported by the University of Mannheim Graduate School of Economic and Social Sciences (GESS). We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Hilal Tanyas, University of Mannheim, School of Social Sciences, Department of Psychology, Mannheim, Germany. Email address: hilal.tanyas@uni-mannheim.de

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, 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. Drift rate, the amount of information needed to decide, and nondecision time showed differential effects between the test formats. Importantly, the item drift rate was slower when item decision was followed by a test for source memory than when item and source were tested in separate blocks, suggesting an earlier onset of source processing during the preceding item test of the standard format.

Keywords: source monitoring, source memory, item memory, diffusion model, temporal sequence

Information Accumulation on the Item Versus Source Test of Source Monitoring: Insights From Diffusion Modeling

Do we first remember an 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 (cf. Reed, 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), but Johnson et al. 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 still 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 Figure 1B in Johnson et al., 1993).

Interestingly, this possible alternative of parallel processing 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 this method whether source information is retrieved serial to completed item processing or already started in parallel to item processing. This is because the response-signal procedure only taps into when source processing is finished, allowing a response, but cannot indicate when source processing started and whether it was already ongoing 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. Albeit the response-lag technique being an insightful temporal study design, we 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 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., upon "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 et al., 2018). 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. 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 (DM; 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 Figure 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 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 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; Ratcliff et al., 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 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 (cf. Osth et al., 2018), 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 measures 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 specifically. 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 versus 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/cqe75/?view_only=a5dbefd433b04dd9aed198b74222b810. The experiment script and the results (including the supplementary analyses) are openly available from https://osf.io/j9zwr/?view_only=da66906fd02a4650992eccbd8171926c.

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 ($\alpha = .05$, 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 were native fluency in English (learned before the age of 6); age (18 –30); normal or corrected-to-normal vision; no diagnosis of mild cognitive impairment; no mental illness daily impact; no head injury caused a knock-out for a period of time; no severe respiratory diseases (i.e., pneumonia or COPD); no medically diagnosed coronary artery or heart issues; no use of medication affecting cognition. 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 CSIMs; 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.co/; 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, the diffusion model parameter estimates could not be obtained for one participant. Thus, the results reported are based on 59 participants (34 female, 24 male, 1 preferred not to indicate sex; M_{age}= 25.51 years, age range = 19 –30). The experiment lasted approximately 30 min. Participants received payment according to the Prolific-set rate of 6£/hour.

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 Supplementary Material).

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 turned out not eligible to participate, the session was terminated, and they received partial payment.

Eligible participants continued with to 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-point (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 Figure 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 Figure 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 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” standing for “TOP” and “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 of “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

We restricted our analyses to only targets because there cannot be source memory for distractors as they were never presented with a source. The thresholds of the diffusion model were linked to response accuracy. Therefore, in the item test, the upper and lower thresholds stand for correct and incorrect target detections (i.e., hits and misses, respectively). However, 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 4000 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 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., 2018). Prior to all analyses, we thus excluded a total of 3.19 % of trials across participants.

Using the software fast-dm (Version 30.2; Voss & Voss, 2007; Voss et al., 2015), we estimated parameters separately for each participant. Drift rate (v), threshold separation (a), and nondecision time (t_0) were free to vary as a function of memory type (i.e., item vs. source), resulting in two drift rate, two threshold separation, and two nondecision time estimates per participant. Since the thresholds are associated with accuracy, there should not be a priori bias toward either response (cf. Voss et al., 2013). The relative starting point ($z_r = z/a$) was thus assumed to be unbiased and fixed at .5. Due to the restricted trial numbers, we kept the model as simple as possible and set the intertrial variabilities of drift rate (s_v), starting point (s_{zr}), and nondecision time (s_{t0}) to zero (cf. Lerche et al., 2017). In total, we estimated six parameters per participant. One individual model had to be removed because parameter estimates could not be obtained for that participant. We report the following analyses based on the remaining 59 participants.

An acceptable model fit is a prerequisite for analyzing and interpreting the diffusion model parameters. Figure A1 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 and several RT quantiles of the behavioral data against the corresponding statistics predicted by the diffusion model based on the parameter estimates. The empirical and predicted values are described along *x*- and *y*-axis, respectively. Each data point thus reflects one participant, and the discrete symbols refer to the item and source tests. 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 59 participants included in the analyses. Mean accuracy and RTs are given in Figure 3. 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, $F < 1$, nor memory type, $F(1, 57) = 1.21$, $p = .276$, $\eta_p^2 = .02$, nor their interaction, $F < 1$, was significant. 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 Supplementary Material.

In the analysis of correct RTs, the main effect of memory type, $F(1, 57) = 35.42$, $p < .001$, $\eta_p^2 = .38$, and the test format × memory type interaction, $F(1, 57) = 121.00$, $p < .001$, $\eta_p^2 = .68$, were significant, but not the main effect of test format, $F < 1$. Relevant to our interest, the simple main effect analyses following up on the significant interaction revealed that correct RTs in the item test of the standard format were slower than in the item test of the blocked format, $F(1, 57) = 17.62$, $p < .001$, $\eta_p^2 = .24$. In contrast, correct RTs in the source test of the standard format were faster than in the source test of the blocked format, $F(1, 57) = 55.55$, $p < .001$, $\eta_p^2 = .49$.

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 non-decisional aspects of the test responses (e.g., the motoric response).

Analyses of Model Parameters

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. Mean estimates of the diffusion model parameters for conditions are presented in Table 1.

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 correct response (i.e., upper threshold) in the majority 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 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)[1]. We thus submitted the absolute values of drift rates to the 2 × 2 ANOVA. Neither the main effects of test format, $F < 1$, nor memory type, $F < 1$, was significant. Importantly, test format interacted with memory type, $F(1, 57) = 10.71$, $p = .002$, $\eta_p^2 = .16$ (see Figure 4). Follow-up simple main effects analyses[2] revealed that during the item test, participants accumulated information slower in the standard format than in the blocked format, $F(1, 57) = 5.32$, $p = .025$, $\eta_p^2 = .09$. This pattern is in line with our preregistered H2 supporting temporal overlap of item and source processing in the standard format. This transfer of source processing to the item part in the standard format is further supported by the descriptively reversed pattern in source decision speed which in the standard format was descriptively faster than in the blocked format. However, this difference did not reach statistical significance,[3] $F < 1$. In the following, we report additional exploratory analyses on the other diffusion model parameters.

Threshold separation (or boundary separation; parameter *a*) is informative for how much information is required to decide for a response. For threshold separation, there were no significant main effects of test format, $F < 1$, or memory type, $F(1, 57) = 3.01$, $p = .088$, $\eta_p^2 = .05$, but the test format × memory type interaction was again significant, $F(1,$

57) = 53.83, p < .001, $\eta_p^2 = .49$. We again performed simple main effects analyses upon the interaction. Larger amount of information was needed for an item response in the standard format compared to the blocked format F(1, 57) = 10.66, p = .002, $\eta_p^2 = .16$. As threshold separation is part of the item and source decision making of interest to us, this pattern further supports a transfer of source processing to the item test in the standard format. Notably, and further in line with temporal overlap of item and source processing, 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, 57) = 9.22, p = .004, $\eta_p^2 = .14$.

Nondecision time (parameter t_0) estimates the remaining time outside the diffusion process such as encoding stimulus and execution of the motor response. It is thus not informative about item and source processing per se but of differences in these additional demands between the two test formats. For nondecision time, the main effects of test format, F(1, 57) = 9.62, p = .003, $\eta_p^2 = .14$, and memory type, F(1, 57) = 73.00, p < .001, $\eta_p^2 = .56$, as well as their interaction, F(1, 57) = 69.57, p < .001, $\eta_p^2 = .55$, were significant. The simple main effects analyses further showed that during the item test, extradecisional processes took significantly longer in the standard format than in the blocked format, F(1, 57) = 4.66, p = .035, $\eta_p^2 = .08$. In contrast, during the source test, participants in the standard format had shorter nondecisional time compared to those in the blocked format, F(1, 57) = 76.42, p < .001, $\eta_p^2 = .57$. Thus, item responses were slowed down in the standard format not only due to slowing of item processing (drift rate, 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. The speed of information accumulation in the item test differed by test format such that accumulation was slower in the standard format compared to the blocked format. This suggests that participants knowing that they would be tested with an ensuing source test next upon their item decisions (i.e., an “old” response) commenced their source processing already during the item test of the standard format, thus leading to a slower speed of information accumulation at the item test stage compared to that of the blocked format. This was further supported by a larger threshold separation for item decisions in the standard compared to the blocked format and descriptive speed up of information accumulation in the source test of the standard format (not significant for accumulation but significant for threshold).

Given our preregistered hypotheses, our results showed convergent evidence for the *possibility* of temporal overlap between item and source information (cf. Tanyas & Kuhlmann, 2023). Accordingly, it seems reasonable to tentatively infer that participants retrieved source information in parallel to item information to some extent when they were tested for their item memory and that this parallel retrieval led to a cost in information accumulation in the item test. However, we additionally address an alternative explanation.

More specifically, the start of source decisions in the item test may not necessarily mean that item and source processing intertwine. Instead, the onset of source processing might have started after the end of item processing but the item response was held off. Therefore, item and source processing may also be happening on the same item test stage, indicating a temporal overlap regarding stage, but without parallel processing of both. In sum, our results underline the transfer of information accumulation from the source test of the standard format to its item test, represented by lower item drift rates in the standard format compared to the blocked format. Yet, they cannot discard alternative time-course scenarios regarding details of this temporal overlap, that is, whether the item and source processing is indeed conducted in parallel or strictly sequential but overlapping in the item test stage. Besides, conceiving 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. An endeavor that we invite future research is to examine factors that could have a natural influence on the degree of overlap (e.g., aging when considered the differential deficits on item and source memory [cf. Old & Naveh-Benjamin, 2008]) or manipulations that could result in some changes on the retrieval course of item and source information (e.g., via different encoding conditions [cf. Lindsay & Johnson, 1991] or different test designs [cf. Fox & Osth, 2022]).

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), it is not yet well understood which component(s) of memory (i.e., encoding, maintenance [vs. forgetting], retrieval) “the quality” maps on. Separating these components and examining their corresponding properties with the drift rates remains as an important challenge for the future.

Apart from the drift rates, we also 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 being already in the item recognition state requires less information to decide for a source response (cf. Tanyas & Kuhlmann, 2023).

Overall, the distinct patterns both in the drift rate and threshold separation suggests that our test format manipulation did really affect the decisional components underlying the item and source responses. However, further analyses on nondecision time showed that the test format manipulation not only yielded differences in item and source decisions but additionally affected nondecision-related factors. For example, comparisons with the blocked format indicated that in the standard format, nondecision time increased during the item test but 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, 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. Yet, given the good fit of the model to the current data, our study provides additional support to other studies showing that the diffusion model can offer reliable account of the data even with small trial numbers (Lerche et al., 2017; Lerche, Neubauer, et al., 2018). 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., 2018). 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., 2018). We recommend episodic memory researchers to consider the feasibility and benefit of the diffusion modeling approach, especially in RT studies.

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 response. For the item response, we found slower speed of information accumulation, need for more information to decide, and increased nondecision time, resulting in a slower response in the standard format compared to the blocked format. The source decision speeds did not change considerably across the test formats. 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, the slower speed of information accumulation in the item decision of the standard format suggests that source processing started during the item test stage. Thus, it is not warranted to assume a sharp separation between item and source retrieval even though they were probed in that order by the standard testing. Instead, there is some overlap of item and source processing on the standard source-monitoring test such that the serially following source response is already prepared in the item test stage.

Declarations

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant KU3329/1-1, project number 392884168. Hilal Tanyas was further supported by the University of Mannheim Graduate School of Economic and Social Sciences (GESS).

Conflicts of interest

The authors report no conflict of interest.

Ethics approval

All procedures performed in the study were in accordance with the principles of the Declaration of Helsinki, the guidelines of the German Psychological Society (DGPs), and the guidelines of the University of Mannheim ethics committee.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent to publish

The authors affirm that participants provided informed consent for publication of the anonymized data.

Availability of data and materials

The datasets analyzed during the current study are available in the OSF repository and can be accessed via the at https://osf.io/j9zwr/?view_only=da66906fd02a4650992eccbd8171926c.

Code availability

The analysis codes and experiment scripts can be accessed via the OSF at https://osf.io/j9zwr/?view_only=da66906fd02a4650992eccbd8171926c.

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

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

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

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

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

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

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

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

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

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* [Unpublished doctoral dissertation]. New York University.

Lerche, V., Christmann, U., & Voss, A. (2018). 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. (2018). 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*, Article 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

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. (2018). 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

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., 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

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

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., 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

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.

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., 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

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. (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., 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*, Article 336. https://doi.org/10.3389/fpsyg.2015.00336

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

Footnotes

[1] 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, accuracy in this study was assessed within the scope of source-monitoring perspective. However, these were lenient criteria, and in our included dataset, there were still 10 participants whose item and/or source drift rates were negative, which indicates that their decision usually ended on the incorrect response (i.e., lower 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' answer were not guessing-free either, but an exclusion of those 10 participants can be seen as a proxy to memory. Findings yielded similar patterns. The main effects of test format and memory type were not significant (*F*s < 1), but their interaction was significant, $F(1, 47) = 9.30$, $p = .004$, $\eta_p^2 = .17$. The simple main effect analyses revealed that information accumulation in the item test of the standard format ($M = 0.49$, $SE = 0.08$) was slower than in the item test of the blocked format ($M = 0.83$, $SE = 0.14$), $F(1, 47) = 4.08$, $p = .049$, $\eta_p^2 = .08$. However, the source decision speeds in the standard ($M = 0.65$, $SE = 0.12$) and in the blocked format ($M = 0.52$, $SE = 0.08$) did not differ significantly, $F < 1$.

[2] An alternative way to probe this interaction (also valid for the other parameters) is the simple main effects of memory type at the test formats (see our *R* code in the OSF for these analyses). However, as preregistered, here we prefer to compare the same (item vs. source) test in the different formats since they were the identical judgments across conditions and thus seem most appropriate for meaningful comparison interpretation.

[3] Note that because the source test was conditional on the item test (i.e., an “old” response), the source trials entering the analyses were fewer than the item trials, indicating that there was comparatively less power to detect the differences in the source test.

Table 1

Group Mean Estimates of the Diffusion Model Parameters

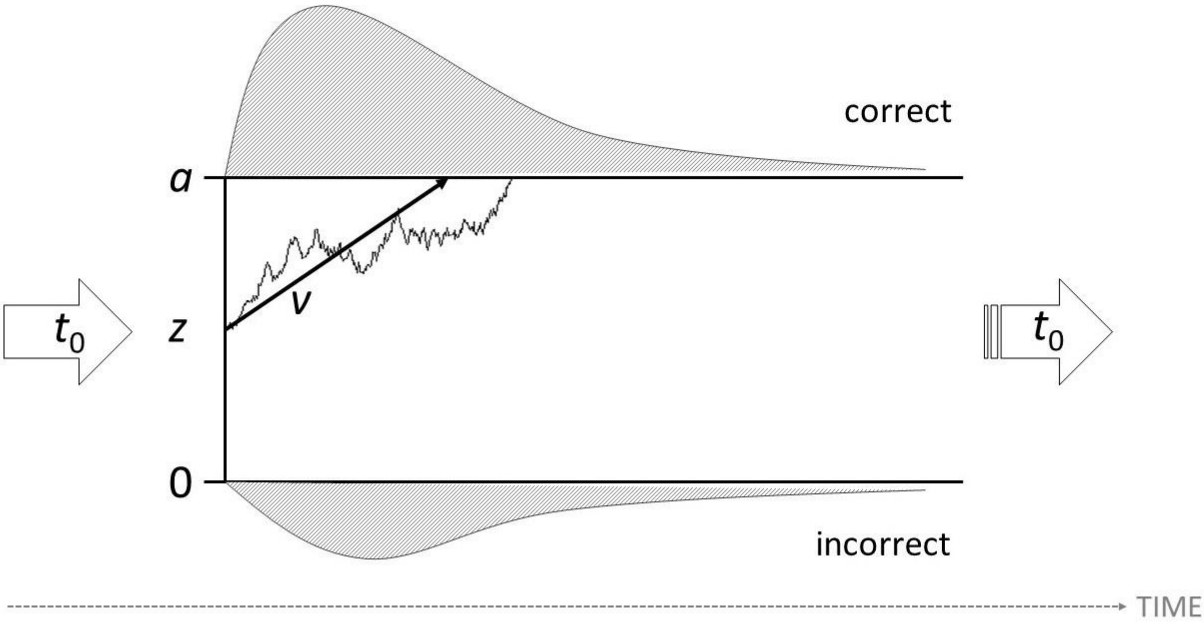
		Test format	
Parameters	Memory type	Standard format	Blocked format
Drift rate (v)[a]			
	v_{item}	0.42 (0.07)	0.77 (0.13)
	v_{source}	0.64 (0.11)	0.51 (0.07)
Threshold separation (a)			
	a_{item}	1.64 (0.08)	1.33 (0.05)
	a_{source}	1.26 (0.07)	1.57 (0.07)
Nondecision time (t_0)			
	$t_{0\text{ item}}$	0.55 (0.04)	0.45 (0.02)
	$t_{0\text{ source}}$	0.15 (0.01)	0.44 (0.03)

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.

Figure 1

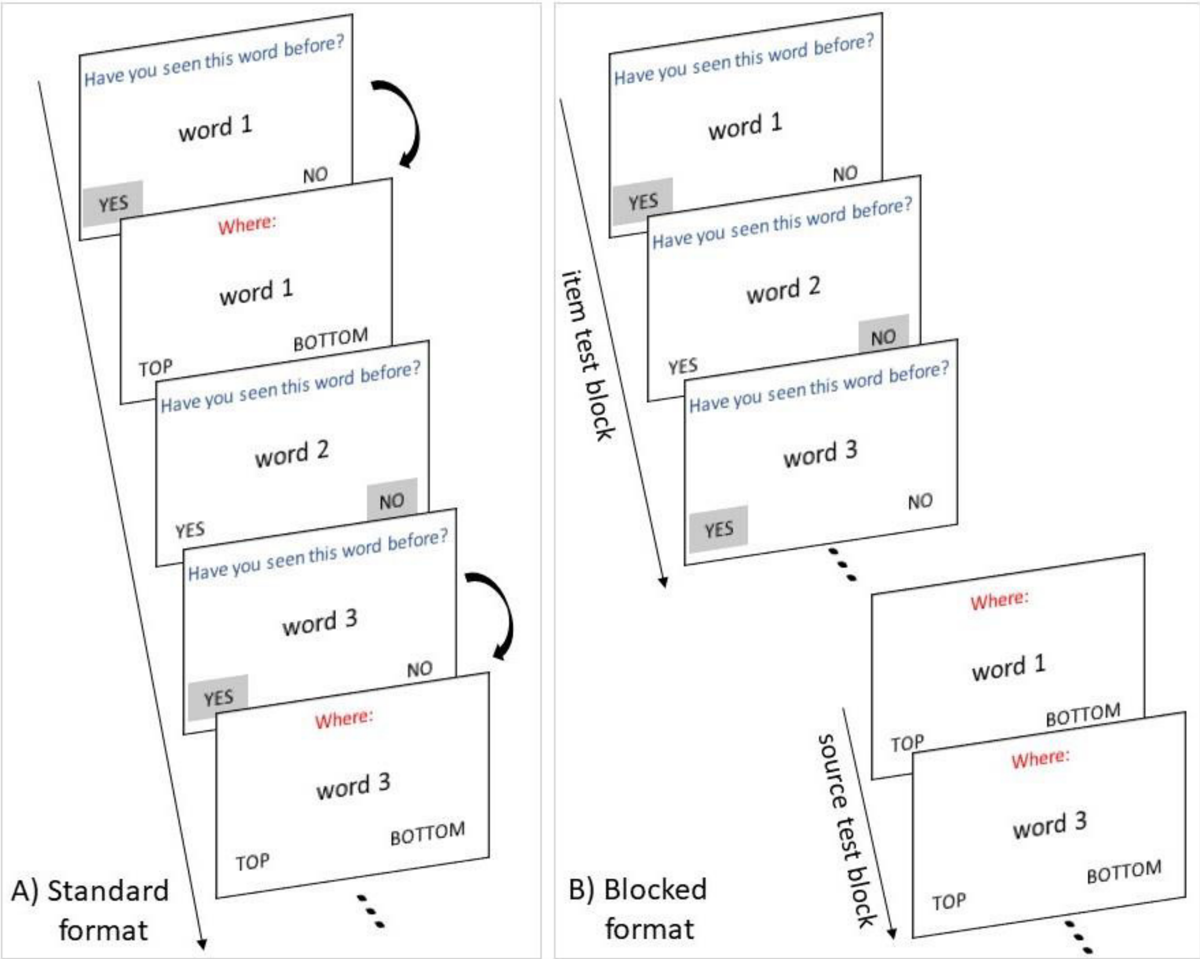
An Illustration of the Diffusion Model



Note. This figure illustrates the decision process as proposed by the diffusion model (Ratcliff, 1978). Here, the upper and lower thresholds correspond to correct and incorrect responses, respectively. 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 the correct (incorrect) response is shown above (below) the respective threshold.

Figure 2

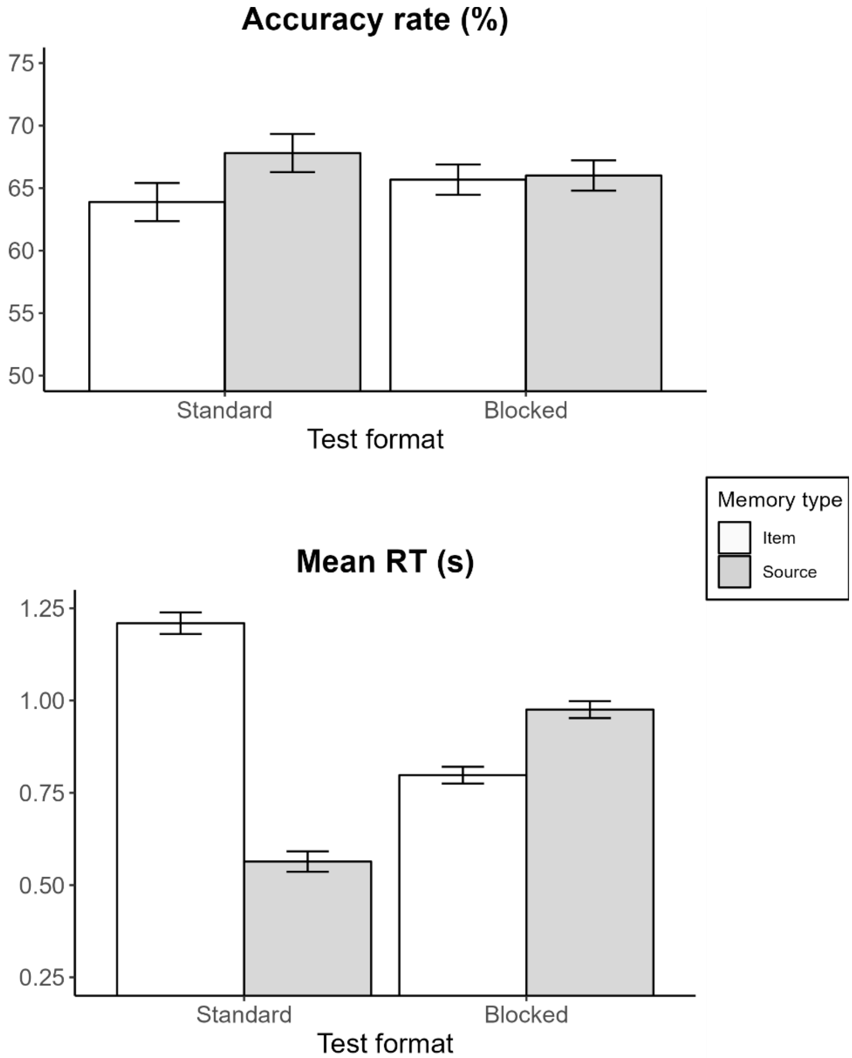
Example Visualizations of the Test Formats



Note. 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.

Figure 3

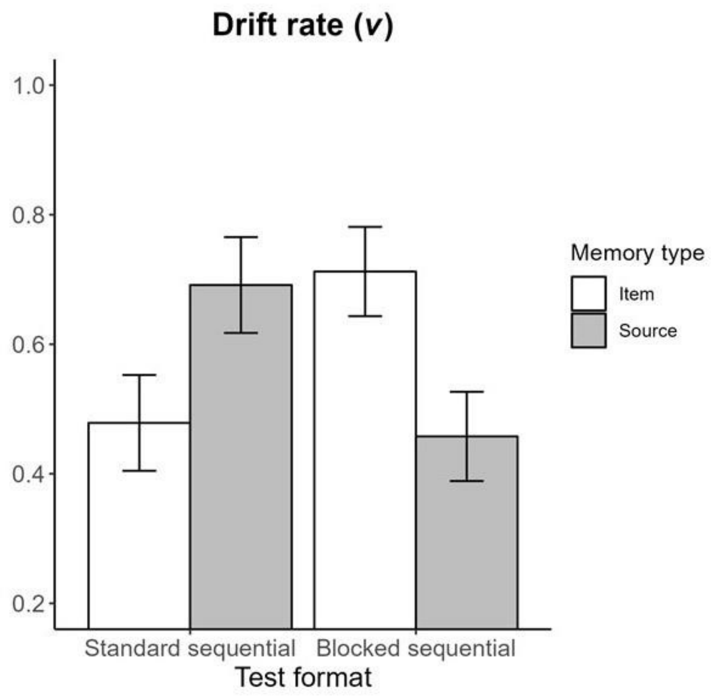
Empirically Observed Mean Accuracy Rate and Correct RTs Across Conditions



Note. Error bars represent standard errors. Mean RT = Mean response time of correct responses.

Figure 4

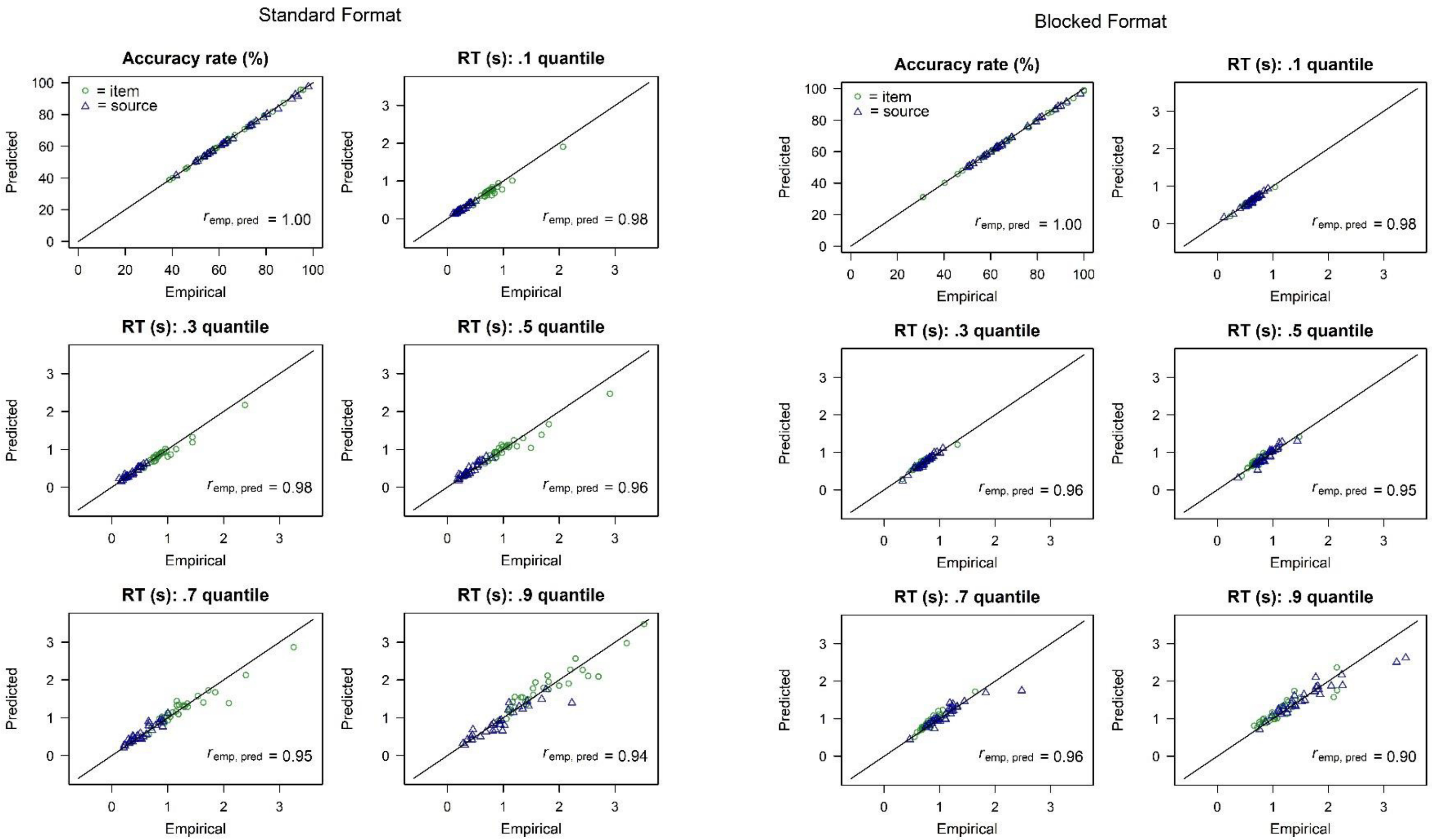
Absolute Drift Rates Across Conditions



Note. Error bars represent standard errors.

Figure A1

Graphical Displays of Model Fit



Note. 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; $r_{\text{emp, pred}}$ = correlation between empirical and predicted statistics.

Supplementary Material

We measured item memory, source memory, and guessing biases separately with the two-high-threshold multinomial processing tree model of source monitoring (2HTSM; Bayen et al., 1996). Based on response frequencies across participants for each item type, the parameters of the 2HTSM represent latent cognitive processes with the probability estimates (see Erdfelder et al., 2009, for a general overview of MPT models). The specific submodel that we used here (see Bayen et al., 1996, for a detailed overview of alternative model versions) describes source-monitoring processes with four parameters: D (item memory with the assumption of equal detection of items presented on the top or bottom and new distractor items), d (source memory with the assumption of equal probability of remembering the top or bottom source), b (probability of guessing that an item is old), g (probability of guessing source A, assuming equal source guessing when source memory fails independent of item recognition status). Using the *multiTree* program (Moshagen, 2010), we fit a joint MPT model, which estimates source-monitoring processes in each test format and allows for their comparisons across conditions. We assessed model fit via maximum likelihood estimation methods and the G^2 statistic. For test of parameter differences across conditions, we relied on the chi-square difference test statistic ΔG^2. The most basic four-parameter submodel of the 2HTSM fit the data well, $G^2(4) = 5.91$, $p = .206$. Table S1 shows parameter estimates by test format. Next, we restricted all four parameters to be equal across the two conditions to test the effect of test format. Note that we did not indicate any specific hypotheses about the parameter differences across the test formats, so here we rather investigated them as exploratory.

Restricting parameter D (i.e., item memory) significantly decreased the model fit, $\Delta G^2(1) = 6.20$, $p = .013$, indicating that item memory differed significantly by test format. We observed that item memory was higher in the blocked format than in the standard format (cf. Tanyas & Kuhlmann, 2023). The immediate source test might have altered the characteristics of old-new recognition accuracy in the standard format (cf., Mulligan et al., 2010). The model fit also became marginally worse upon restrictions on parameter *b* (i.e., item guessing), $\Delta G^2(1) = 3.63$, $p = .057$. In the blocked format, it is not surprising that participants were more lenient to guess “old” upon no item detection since they were not informed about the upcoming source test block (i.e., an “old” response did not immediately have the consequence of further query into memory). Restrictions on parameter *d* (i.e., source memory), $\Delta G^2(1) = 0.16$, $p = .685$, and parameter *g* (i.e., source guessing), $\Delta G^2(1) = 0.26$, $p = .608$, did not significantly decrease the model fit, implying that neither source memory nor source guessing differed significantly by test format. Source guessing averaged across the test formats, $g = .52$, 95% CI [.50, .55], indicates a significant bias to guess top, $\Delta G^2(1) = 4.02$, $p = .045$, but this tendency was comparable across the conditions.

Supplementary References

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

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Mulligan, N. W., Besken, M., & Peterson, D. (2010). Remember-know and source memory instructions can qualitatively change old-new recognition accuracy: The modality-match effect in recognition memory. *Journal of Experimental Psychology, 36*(2), 558–566. https://doi.org/10.1037/a0018408

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

Table S1

Parameter Estimates and Confidence Intervals for Different Conditions of Test Format

	Model parameters			
Test format	*D*	*b*	*d*	*g*
Standard format	.38 [.34, .41]	.38 [.35, .41]	.57 [.47, .66]	.52 [.48, .55]
Blocked format	.44 [.40, .47]	.42 [.39, .45]	.59 [.51, .67]	.53 [.50, .56]

Note. The presented model parameters are probability estimates that can range from 0 to 1. *D* = item memory; *b* = item guessing (chance level is .5); *d* = source memory; *g* = source guessing (estimates higher than the chance level of .5 indicate guessing bias towards “top”; estimates lower than .5 indicate guessing bias towards “bottom”). Brackets indicate 95% confidence intervals.