Memory & Cognition
https://doi.org/10.3758/s13421-026-01930-1

Dissociable effects of schema congruency on item recognition and source memory across levels of expectedness and surprise

Amnon Yacoby[1] · Andrea Greve[2,3] · Erin Schaverien[1] · Limor Shtoots[4] · Richard N. Henson[2] · Daniel A. Levy[1,5]

Received: 11 December 2025 / Accepted: 24 July 2026

Abstract

Prior knowledge can enhance memory for both schema-congruent and schema-incongruent events. According to the SLIMM framework, memory performance generally follows a U-shaped pattern, with both highly expected and highly surprising events being better remembered than neutral ones. We tested this prediction across four experiments in which participants viewed scenes containing congruent or incongruent objects, rating each as expected or surprising on a four-level scale. Memory was assessed using: (1) old/new item recognition categorized by recollection and familiarity, and (2) source memory, in which participants reported whether objects had been rated as expected or surprising at study. Schema effects on memory varied by test type. Item recognition consistently showed an advantage for expected trials only, primarily via recollection, while familiarity (if assumed independent of recollection) increased with surprise level—contrary to SLIMM's predictions. We hypothesize that schema-related retrieval strategies enhance recollection of expected objects. Source memory did produce a reliable U-shaped pattern, even though this particular task should only produce an advantage for surprising trials according to SLIMM. We suggest participants tagged events as "surprising" or "expected" during study and could later recall the tag. When the tag was 'weak' or unavailable—more likely for moderately categorized events—participants fell back on schema-related retrieval strategy, which benefits retrieval of expected events. Together, these mechanisms produce the observed U-shaped pattern for source memory. Our findings reveal the complex retrieval test dependent manner in which prior knowledge shapes episodic memory.

Keywords Surprise · Congruency · Schema · Retrieval · Memory

Introduction

In everyday life, most objects we encounter align with our prior knowledge of specific environments. For instance, objects commonly found in kitchens, such as kettles, plates and dishwashers, conform to an environmental "schema" (e.g., Anderson, 1983; Bartlett, 1932; Rumelhart, 1980). When we encounter new information consistent with this schema, such as a new pot in the kitchen, it activates relevant prior knowledge, facilitating the comprehension and integration of this new information (Alba & Hasher, 1983). Conversely, when we encounter items that are incongruent with the schema, such as a bed in the kitchen, it induces surprise (Itti & Baldi, 2009; Lorini & Castelfranchi, 2006; Schank, 1986), potentially facilitating the formation of a memory of the surprising event.

✉ Amnon Yacoby
amnon.yacoby@gmail.com

Andrea Greve
andrea.greve@mrc-cbu.cam.ac.uk

Erin Schaverien
erinschaverien15@gmail.com

Limor Shtoots
slimor@runi.ac.il

Richard N. Henson
Rik.Henson@mrc-cbu.cam.ac.uk

Daniel A. Levy
daniel.levy@runi.ac.il

1 School of Psychology, Reichman University, 8 HaUniversita St., 4610101 Herzliya, Israel

2 Medical Research Council Cognition and Brain sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, UK

3 Department of Psychology, University of Cambridge, Cambridge, UK

4 School of Medicine, Reichman University, 8 HaUniversita St., 4610101 Herzliya, Israel

5 Department of Psychology, Palo Alto University, Palo Alto, CA, USA

A substantial body of research highlights both the mnemonic advantages of expected (schema-congruent) events and surprising (schema-incongruent) events. In general, the degree of surprise or expectedness can vary across people, depending on their past experiences (i.e., different schemas/priors). Thus, we will use the terms "congruent/incongruent" to refer to experimentally manipulated conditions, and "expected/surprised" to refer to the subjective experience they induce. Congruency benefits have been reported across various domains, including scene-item associations (Liu et al., 2018; van Kesteren et al., 2013a), narrative comprehension (Bransford & Johnson, 1972) and item–color pairings (Maril et al., 2011). For example, Staresina et al. (2009) asked participants to evaluate word-color pairs using two types of judgments: "plausibility" (conceptual coherence of pairs; e.g., "gray elephant") and "valence" (aesthetic appeal of pairs: e.g., "blue apple"). It was found that congruency improved subsequent memory for both the specific word pairing and for the encoding context (judgment type).

On the other hand, a large body of research has shown that incongruent or surprising events are remembered better than neutral ones (Bein et al., 2021; Greve et al., 2017; Hunt & Worthen, 2006; Tulving & Kroll, 1995). A classic example is the "isolation effect" (or "Von Restorff effect"), in which a distinct item (e.g., "potato" in a list of fruits) is more likely to be recalled due to its distinctiveness (S. R. Schmidt, 1991; Von Restorff, 1933). Such advantages have been found in item recognition (Erdfelder & Bredenkamp, 1998) and in source memory tests (Küppers & Bayen, 2014). One account of this mnemonic advantage appeals to the distinctiveness of schema-incongruent information at retrieval (McDaniel et al., 1995), while another account appeals to increased attention during the encoding of surprising stimuli (the "attentional elaboration" hypothesis; Brewer & Treyens 1981; Friedman 1979). A related account assumes that prediction error causes better encoding of surprising events (Henson & Gagnepain, 2010; Quent et al., 2021), possibly through triggering such increased attention. Yet another account, the copy-and-tag account, argues that schema-inconsistent information is facilitated and stored separately (with a "tag"), whereas schema-consistent information is integrated into the schema at a shallow, gist-based level (Graesser et al., 1979; Graesser & Nakamura, 1982).

When combining these mnemonic effects, the relationship between schema-congruent and incongruent events is expected to be nonlinear: Both highly congruent and highly incongruent events should be remembered better than moderately congruent or incongruent events. This "U-shape" pattern is predicted by the schema-linked interactions between medial prefrontal and medial temporal lobes (SLIMM) framework (van Kesteren et al., 2012), which integrates schema theory with recent neuroscientific findings.

According to SLIMM, schema-congruent information is supported by the medial prefrontal cortex (mPFC) enabling rapid integration of event information into cortical representations (Tse et al., 2007). In contrast, schema-incongruent events rely more on the medial temporal lobe (MTL), particularly the hippocampus, to encode unexpected occurrences (Eichenbaum et al., 2007). Thus, the two brain systems operate at opposite ends of the U-shaped function, each improving memory, but through dissociable mechanisms. The congruency advantage is thought to be driven by easier or better assimilation of information either at study or during consolidation (Alba & Hasher, 1983; Bransford & Johnson, 1972; Tse et al., 2007), whereas the incongruency-advantage is attributed to high prediction error that enhances encoding (Quent et al., 2021).

SLIMM further assumes that surprising occurrences trigger storage of an episodic snapshot of the whole event, including other contextual information. The rationale is that such contextual information may be important for updating schema in the future so that the same prediction error does not recur. As a consequence, SLIMM predicts that memory for surprising items will be associated with recollection (Yonelinas, 2002; Yonelinas et al., 2024). Memory for expected items, on the other hand, is more likely based on familiarity (due to the integration with a schema that specifically discounts contextual information).

Several studies have explored episodic memory performance across a range of congruency levels. For example, van Kesteren et al. (2013a) paired scenes with objects that were congruent (e.g., classroom–chalk), neutral (taxi rank–pen), or incongruent (tennis court–soup ladle). Memory increased linearly with congruency, such that item recognition and associative memory were best for congruent, followed by neutral, and worst for incongruent trials. In other words, they did not observe the U-shape predicted by SLIMM. Similarly, Frank et al. (2018) reported better memory for schema-congruent images, and worse memory for incongruent ones, relative to neutral ones. One possible explanation for the lack of any incongruency advantage in these studies is the strategic use of schema at retrieval. When presented with a single object, participants may generate semantic associates that cue the studied episode ("generate-and-recognize"; Watkins & Gardiner, 1979); a strategy that can only be successful for congruent trials. For example, thinking of semantic associates of the "chalk" cue may have led to retrieval of a "classroom" in the van Kesteren et al. (2013a) study. Such generation of schema-related scene associations would disproportionately benefit schema-congruent items, potentially masking mnemonic advantages for unexpected items.

Some studies have minimized this "schema-related retrieval" confound. For example, Greve et al. (2019) taught arbitrary rules to create expected, neutral and surprising events, and then tested associative memory for these events. Importantly, the events could not be uniquely generated by the rule alone. The ineffectiveness of a schema-related retrieval strategy may be why these authors found advantage of schema (rule) incongruent events as well as congruent ones, relative to neutral ones, consistent with the U-shape predictions of SLIMM.

Other studies have controlled for schema-related retrieval but used preexperimental knowledge rather than training new schema. Using immersive virtual reality for example, Quent et al. (2022) varied the congruency of the locations of everyday objects within a kitchen. Importantly, memory was tested by forced choice in which the foil locations were matched for congruency, which again revealed the U-shape predicted by SLIMM. However, remember/know judgments showed that memory for both expected and unexpected locations was associated with recollection, contrary to SLIMM (which predicts familiarity rather than recollection for expected items). Ortiz-Tudela et al. (2024) examined memory across ages (children, young adults, and older adults) using cartoon strips that ended in different ways. Importantly, one of their memory tests was to ask about the nature of the ending (i.e., whether it had been expected, neutral, or unexpected). Performance on this "source memory" task was broadly U-shaped, but its profile varied by age: younger groups showed better source memory for unexpected endings, while older groups showed better source memory for expected endings. Memory for the contents of the cartoons ("item memory"), however, yielded a different pattern of results, further indicating that retrieval demands might modulate the effects of expectedness/surprise.

Taken together, previous evidence showed that when schema-related retrieval bias is controlled, a U-shaped pattern can be observed, at least for associative memory and source memory, even if the situation for item memory is less clear. Moreover, although this U-shape pattern is consistent with SLIMM, the advantage for expected trials has been associated with recollection, not familiarity, contrary to SLIMM's predictions. More generally, the pattern of results seems to depend on the type of retrieval test (van Kesteren et al., 2012) and/or the retrieval strategy employed by participants (Brainerd et al., 2025).

To reconcile these findings and further test the predictions of SLIMM's and other frameworks, we examined how memory tests and retrieval goals shape performance across a continuum of participant-specific expectedness and surprise. We used object–scene pairs, like van Kesteren et al. (2013a), but with congruent objects (e.g., a tennis ball) or incongruent objects (e.g., a bear) that were fully integrated into a scene (e.g., a tennis court), rather than simply shown adjacent to each other. Such integration is more similar to ecological perceptual experience (Davenport & Potter, 2004) that could generate a stronger prediction error signal for unexpected objects or lead to greater feeling of congruity than in van Kesteren et al. (2013a), potentially facilitating memory of surprising items. After each trial, participants rated the congruency on a four-point scale, from highly expected to highly surprising, based on their general knowledge. These ratings were used to define the level of expectedness/surprise in the analysis of later memory performance, given that the subjective experience is likely to vary across people. Participants were notified that a memory test would follow the study.

We first tested item recognition, in which participants identified an object as either "old" (studied) or "new" (unstudied). To minimise guessing, a third "don't know" option was offered (Chua et al., 2012; Roediger et al., 2022). Importantly, new objects (foils) were congruent with one of the studied scenes, so that generating a congruent scene was not sufficient to determine whether an object had been studied. For those items correctly called "old," participants were further asked whether the item was "remembered" or "known," to assess recollection and familiarity (Gardiner & Richardson-Klavehn, 2000; Yonelinas et al., 2024; see Methods).

Next, we assessed source monitoring (Johnson et al., 1993). Participants indicated whether each old objects had appeared in an expected or surprising context, with a "don't know" option to allow for guesses (Bayen et al., 2000; Kuhlmann & Touron, 2012). Crucially, they were told that retrieving the specific scene was unnecessary; only whether it had rendered the object to be judged expected or surprising. Because this task focused participants on recalling only the degree of expectedness (Dobbins & McCarthy, 2008; Ortiz-Tudela et al., 2024), it should be less susceptible to schema-related retrieval and guessing than standard episodic memory tasks. For example, Gray and Gallo (2015) demonstrated that orienting retrieval toward recollection via a recapitulation strategy (Rugg & Wilding, 2000) effectively reduced reliance on familiarity-based strategy in a recognition memory task. In Experiment 2, we added an associative memory test using a four-alternative forced-choice (4AFC) task, which presented four scenes (including a target and a semantically related foil scene).

To summarise, according to SLIMM, in the absence of schema-related retrieval, overall performance in item recognition should show a U-shaped function, with performance at the extremes driven more often by recollection for unexpected items, but by familiarity for expected items. Likewise, associative memory should show a U-shaped function, with better memory for both highly expected and unexpected pairs, relative to moderate pairs. For source memory however, SLIMM predicts memory should be best for surprising trials, since this is when SLIMM claims that contextual details are also stored. However, all these predictions can be overridden if other strategies are used at retrieval, such as using schema to reconstruct memory for an event.

General experimental design

We conducted four experiments (see Table 1). All experiments used the same study phase, as detailed below, but varied in the retrieval methods employed in the test phase. Experiments 1a and 1b addressed our basic research question concerning the putative parallel benefits of surprise and expectedness on memory performance (i.e., the U-shaped pattern) by employing a variety of memory tests. In Experiment 1a, retrieval occurred shortly after encoding, while in Experiment 1b, retrieval occurred after a delay of approximately 24 hours, to see if delay moderated or dissociated effects of surprise or expectedness (e.g., if consolidation processes were biased towards schema-congruent information, as suggested by van Kesteren et al., 2012; van Kesteren et al., 2013b). All experiments were conducted within participants.

Methods

Participants

A total of 138 native Hebrew-speaking or native English-speaking university students participated in one of four experiments. See participants numbers and demographic data per experiment in Table 1. All participants had normal or corrected-to-normal vision and were screened to ensure that they had no known psychiatric diagnoses. Participants were compensated with either monetary compensation or university credit. Each participant received verbal instructions, on-screen instructions and feedback in their primary language. Written informed consent was obtained from all participants prior to participation. Consent was obtained for Experiments 1a, 1b, 2b in a manner approved by the Reichman University institutional review board, and for Experiment 2a in a manner approved by the Cambridge Psychological Research Ethics Committee. Participants performing at chance levels were excluded. For item recognition, we excluded participants whose false alarm rate did not significantly differ from 50% for the new items (binomial test). For source memory, we excluded participants whose response distribution did not significantly differ from chance (33.3% of all tested items; per expected/surprising/don't know categories; chi-square goodness-of-fit test). Two participants were excluded from the item recognition analyses (one from Experiment 2a and one from Experiment 2b) and two from the source memory analyses (one each from Experiments 1b and 2b). We used G*Power (Version 3.1.9.7; Faul et al., 2007) to calculate the required sample size for Experiments 1a, 1b and 2a. The key analysis involved computing contrasts and fitting linear mixed-effects models. Previous studies measuring similar U-shaped memory curves have demonstrated an effect size of $d = \sim 0.8$ (Greve et al., 2019). For planning purposes, we assumed a medium effect size ($d = 0.5$) for a matched-pairs t test, with an alpha of 0.05 and a power of 0.80. Based on these parameters, the required sample size was $n = 27$. Expecting some drop-out, we collected data from 27–32 participants in each of these three experiments. Experiment 2b was an fMRI study, and we planned the number of participants with criteria and simulations to match this fMRI study's goal. Based on these criteria, we recruited 47 participants for this study. Following a reviewer's concerns regarding the fit of the power analysis methodology to the planned analyses, we conducted a post hoc simulation-based power analysis to assess whether any of the experiments were underpowered. Results are reported in the Results section.

Stimulus materials

We prepared 200 images, each presenting a typical scene from daily life or nature (e.g., a bakery, parking lot, mountain landscape). Initially, each scene was paired with a schema-congruent object (e.g., a loaf of bread in a bakery), forming a pool of objects across all scenes. From this pool, we then selected three additional objects for each scene, resulting in a total of four associated objects per scene: two schema-congruent objects and two schema-incongruent

Table 1 Experimental conditions

Experiment	Retrieval tests	Other conditions	Number of participants	Mean age [age range]
Experiment 1a	Recognition memory; Source memory		32; 23 females	23.4 (18–27)
Experiment 1b	Recognition memory; Source memory	Retrieval post 24H	32; 21 females	24.2 (18–31)
Experiment 2a	Recognition memory; Source memory; 4AFC		27; 14 females	22.3 (18–35)
Experiment 2b	Recognition memory; Source memory; 4AFC	Encoding in scanner	47; 35 females	24.5 (20–35)

Experiments 2a and 2b employed similar paradigms to Experiment 1a, with the addition of a 4AFC associative memory test. Experiment 2a took place in the lab, whereas in Experiment 2b encoding took place during an fMRI scan objects. However, each scene was presented with only one embedded object per participant.

The two objects that were congruent with the same scene were generally semantically related. Each object–scene pairing was designed so that both objects that are schema-congruent with one scene become schema-incongruent when placed in a specific other scene, and vice versa. For example, for a scene of a classic museum hall we had two congruent objects, a sculpture of a lion and a sculpture of a woman, each embedded into the scene and placed on the hall floor in suitable locations. We then had two incongruent objects, a cigarette packet and an alcohol bottle, each placed on the hall floor. In contrast, these four objects served as incongruent and congruent objects, respectively, when embedded into a convenience store scene (i.e., in the context of a convenient store the sculpture of a lion and sculpture of a woman were incongruent, while cigarette packet and alcohol bottle were congruent items). To determine whether any given scene–object pairing was congruent or incongruent, we conducted an online Qualtrics experiment in which an independent set of participants (who were not part of the main experiments) evaluated 800 combinations (200 scenes, each paired with four objects) and rated the level of expectedness/surprise they felt when observing each of the combinations. Ratings were selected from four categories: "highly expected," "moderately expected," "moderately surprising," or "highly surprising." Each participant rated 50 combinations, evenly split between assumed expected and surprising scene–object pairings, with no scene repeated. Each scene–object combination was rated by 25–30 participants, and the majority rating (above 60%) determined its classification as congruent or incongruent.

Using these ratings, we constructed four sets of stimuli in which each scene was paired with one of the four objects, such that there were 100 assumedly expected scene–object pairs and 100 assumedly surprising scene–object pairs in each set. Each participant was randomly assigned to view one of these sets, with condition counterbalanced across participants. To ensure equal numbers of expected and surprising combinations across all four sets, we adjusted or replaced certain objects and scene images as needed. These modified scene–object combinations were then validated by additional online participant groups. All images of scenes, objects and their combinations are available in the OSF site (https://osf.io/u8fzy/files/osfstorage).

Procedure

Study phase

In each trial of the study phase, participants were instructed to carefully inspect a scene image which they viewed for 4 s. After a short delay marked by a fixation cross, the same scene reappeared on screen with an object embedded within it, remaining visible until participants responded or until 5 s had passed (the latter option was only practiced in Experiment 2b). The embedded object was highlighted by a size-appropriate yellow circle or oval. Participants were instructed to classify the object as highly expected, moderately expected, moderately surprising, or highly surprising within the context of the background scene, using one of four designated buttons on the keyboard. Once a selection was made or 5-s elapsed, the scene–object image disappeared, replaced by a fixation cross. Participants were notified that a memory test would follow the study. See the experimental paradigm schematics in Fig. 1.

A total of 200 scenes were presented, with 100 independently rated as congruent and 100 as incongruent, as previously described. The sequence of trials, both scene-only and scene-plus-object presentations, were mixed with baseline fixation interstimulus intervals (ISI), averaging 1 second and ranging from 0.2 to 1.7 s. All images were presented centrally on a black background, with task instructions shown in white below each image. The 200 scenes and 200 integrated scene–object pairs were divided into four blocks of 50 trials each, with 25 congruent and 25 incongruent pairs per block. Trial order within each block was randomized, and participants had three self-paced short breaks between the four blocks. The study phase lasted about 30 min.

Participants indicated their choice by pressing one of four buttons (*X*, *C*, *N*, *M*) using the middle and index fingers of each hand. Counterbalancing involved each participant being assigned to (a) one of the four stimulus sets (see Stimulus materials above) and (b) one of two orders for the four buttons for the congruency categories judgment (right to left or left to right). By this counterbalancing scheme, each scene appeared in only one condition per participant but equally often in congruent and incongruent conditions across the sample. Before the study phase, participants completed a brief practice session with sample trials.

Test phase

Following the study phase (immediately in Experiments 1a, 2a, 2b, or after approximately 24 hours in Experiment 1b), participants performed two or three types of memory test, in the following order:

Item recognition test Participants viewed 400 single object images: 200 studied and 200 new objects. For each participant, the "new" objects were the 200 items from the 400 object-stimulus set that were not shown during study (see Stimulus materials), such that scene congruency could not determine old/new status. For each object, participants responded whether it was an "old" (previously seen) or "new" (not previously seen) object, or that they "did not

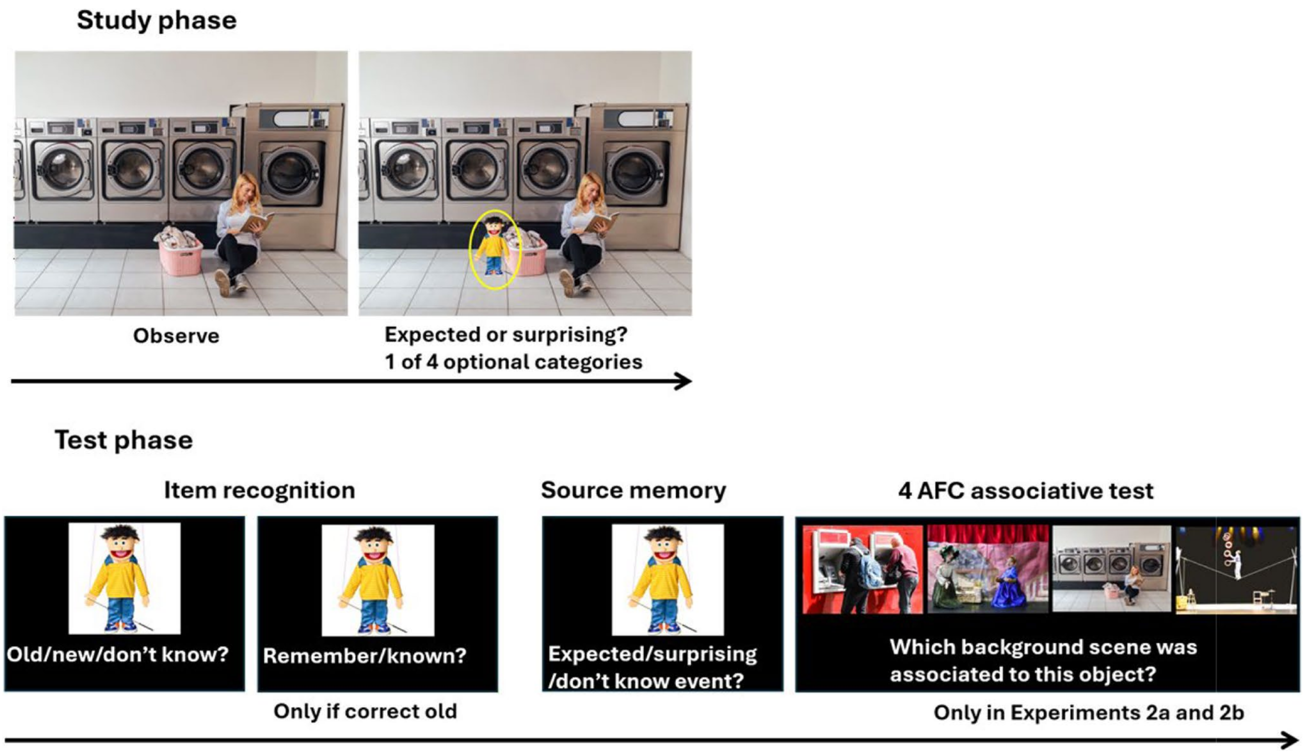

Fig. 1 Experimental Paradigm. *Notes.* During the study phase, participants viewed scene images. Then, an object image was shown integrated in the scene, marked by a yellow circle/oval. The participants' task was to indicate the congruency of the object and scene using 4 levels. Memory was tested immediately after study (Experiment 1a, 2a, 2b) or 24 hours later (Experiment 1b), using an old/new item recognition task with remember/known judgments, followed by a source memory test, in which participants recalled if the event in which the presented object was shown was expected or surprising. In Experiments 2a and 2b, there was also a final 4AFC associative memory test, in which participants selected the background scene in which the object had been originally shown

know." When they recognized an object, they then classified the memory as either "remembered" or "known" (Yonelinas et al., 2024). We instructed participants that they would perform the "remembered" or "known" task only for old objects that they had correctly identified in the preceding "old/new" task. Prior to the recognition test, participants were trained to correctly distinguish between these two types of memory judgments, using instructions based on Migo et al. (2012). After responding, the next trial appeared on screen. There were two self-paced short breaks during this task.

Source memory test Participants were presented with the 200 objects studied earlier and indicated whether each had been judged as expected or surprising when it appeared in the scene image at study. A third "don't know" option was also available. We emphasized that remembering the actual scene was not needed, and that participants should instead focus on retrieving their feeling of expectedness or surprise for the study trial. After a response was made, the next trial appeared on the screen, and a self-paced short break was available after the first 100 trials. Prior to the source memory test, participants received relevant task instructions and training.

AFC associative memory test (Experiments 2a and 2b only) Following the source memory task, participants completed an associative memory test, selecting the specific scene in which the object was embedded during study. On each trial, four scene images were randomly presented in a row on the screen, one target scene and three foils, all of which had been previously studied with other objects. The foils were selected to control congruency effects. If the object was originally part of a congruent scene, one foil was also congruent and the remaining two foils were incongruent or unrelated. If the original pairing was incongruent, one foil was also incongruent and the other two foils were congruent. Thus, each of the 4AFC sets always contained two congruent and two incongruent scenes. Study and test tasks were programmed and administered using PsychoPy (Peirce et al., 2019).

Analyses

We used logistic linear mixed-effects model to fit binary accuracy outcomes for every trial (correct/incorrect) using congruency category (highly expected, moderately expected, moderately surprising, highly surprising) as the independent variable, and participants and items as random intercepts and random slopes. This methodology follows recommendations for confirmatory hypothesis testing aiming to minimize Type 1 errors (Barr et al., 2013). Alternative optimizers were employed where needed to achieve convergence. Near-singularity warnings in some models reflected near-zero variance in stimulus-set random effects, indicating negligible between-stimulus variability; these terms were retained in accordance with Barr et al. (2013).

To assess the critical measures (i.e., potential U-shaped and/or linear trend patterns across the congruency levels), we used a linear + quadratic function to fit the four congruency categories (e.g., Ortiz-Tudela et al., 2023; Quent et al., 2022) using *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages in R (R Core Team, 2021), using the Satterthwaite estimate of degrees of freedom (DF; Satterthwaite, 1946). Given that the full four-level factor was modelled continuously via the polynomial terms, no post hoc contrasts were necessary. Notably, the four congruency levels were based on participants' ratings collected during the study phase.

In the item recognition test in which we separated correct responses of old objects into recollection and familiarity, we assumed independence between recollection and familiarity (Yonelinas, 2002). Accuracy of recollection was measured by assigning 1 to all items called "remembered" and 0 to all other old (studied) items. Accuracy of familiarity was measured by assigning 1 to all items called "known" and 0 to all studied items excluding the remembered items.

The dependent variable is binary, representing the accuracy of test responses (old objects called "old" and new objects called "new"). Although participants could select a "don't know" option to indicate retrieval uncertainty in both item recognition and source memory tests, accurate responses may still contain guesses. We modelled and assessed the guessing component in item recognition and source memory tests using multinomial processing tree (MPT) analysis, as detailed in the Supplemental Material (Erdfelder et al., 2009; Schmidt et al., 2025).

Results

We obtained similar results across all four experiments, so we present summary tables and figures for all experiments together (for each memory test separately), using the congruency categories defined by the individual participants during the study phase.

In the associative memory task, we observed a significant schema bias related to the foil stimuli. This bias was confirmed and corrected in a follow-up Experiment S1. As a result, the outcome of the associative memory task does not inform our central question of how expectedness and surprise influence retrieval. However, for completeness we report the analyses, results and discussion of the associative memory task, including those from Experiment S1, in the Supplemental Material (see narrative and Figs. S8–S9 and Tables S5–S9).

Study phase results

Study task ratings were similar across experiments, with most responses falling in the "highly expected" and "highly surprising" categories. For illustration, Supplemental Fig. S1 shows mean number of responses per congruency category and participant in Experiment 1a. Acceptable response rates exceeded 97% (Table 2). Reaction times (RTs) showed an inverted U-shape across congruency, with slower RTs for moderate categories than for the extremes (see Fig. S2 in the Supplemental Material). Trial-based logistic linear mixed effects model confirmed this inverted U-shape (Table 2). The positive linear component further indicated an additional pattern of slower RTs for "highly surprising" than "highly expected" items. Table 2 also reports the mean and standard

Table 2 RTs for expectedness/surprise ratings at encoding

Experiment	1a	1b	2a	2b
Fitting linear and quadratic model				
Linear comp.	**β=.028; z= 3.48; *p* <.001**	**β=.081; z=9.4; *p* <. 001**	**β=.043; z=5.2; p<.001**	**β=.050; z=−7.5; *p* <.001**
Quadratic comp.	**β=-.55; z=−36.7; *p* <.001**	**β=-.69; z=43.3; *p* <. 001**	**β=-.51; z=−35.7; *p* <.001**	**β=-.58; z=−46.8; *p* <. 001**
DF	6312	6301	5321	8053
Mean RT (SD) - sec	2.21(.52)	2.41(.57)	2.18(.60)	2.46(.56)
Mean response proportions	.989	.989	.990	.976
N Participants	32	32	27	47

Results of linear mixed effects model were fit to study RT in trials across experiments. β and z represent the coefficients of the linear and quadratic components in the models; comp – component; DF is the degrees of freedom. The numbers in parentheses show standard deviations (SD) deviation of RTs (collapsed across congruency) which did not reveal any differences.

Item-recognition test results

Figure 2 shows a consistent congruency advantage across experiments. Recognition accuracy typically decreased from objects rated highly expected to those rated highly surprising (except in Experiment 2b, where highly surprising objects showed a slight increase relative to moderately surprising objects).

Table 3 reports the accuracy analysis results of item recognition per congruency categories. We report the fixed effects results and their confidence intervals, the random effects variances and the degrees of freedom. Logistic linear mixed models fit to trial level item recognition showed a significant negative linear relationship with surprise across all experiments. No significant quadratic component emerged, except for Experiment 2b, where an additional positive quadratic term revealed a left-tilted (declining toward surprising categories) U-shape pattern (Table 3). The R-code model equation is shown below. Additionally, we report in Table 3 the descriptive statistics of item recognition, including the old/new/don't know responses.

Although declared uncertainty ("don't know") rates were low, we used multinomial processing tree (MPT) to model latent guessing within 'accuracy' responses. Optimized model results indicated negligible contributions of guesses ≤2% (see Supplemental Material Table S1 and Figs. S3–S4).

R-code model equation:

$$model.old.items.sq = glmer(old\_items\_accuracy\ \sim congruency + I(congruency^2)+ (1 + congruency|participant) + (1 + congruency|items), data.old, family = binomial)$$

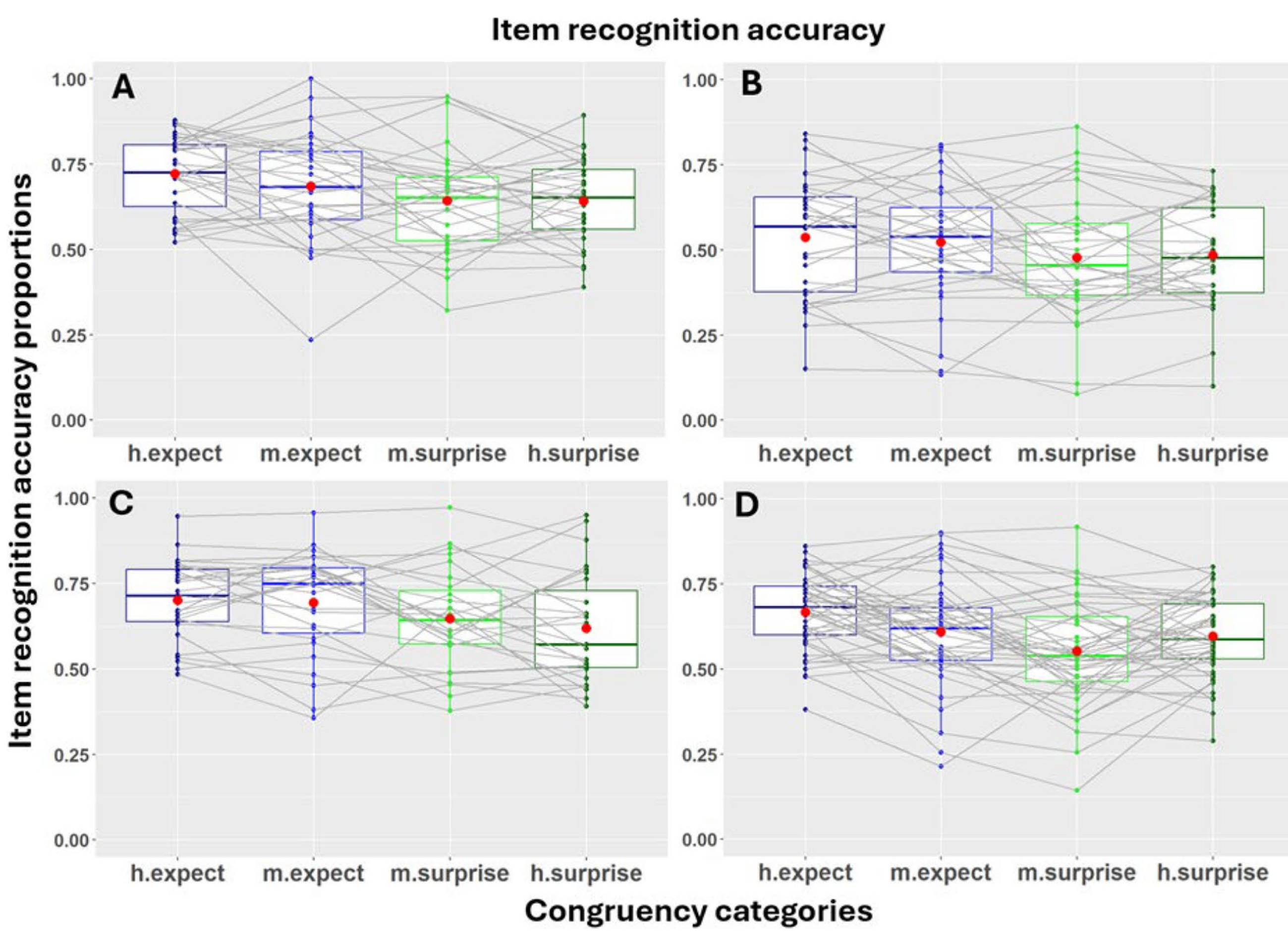

Fig. 2 Item Recognition Accuracy. *Notes.* Boxplots of trial-averaged item recognition accuracy for each participant (joined by lines) for each of the four congruency categories that were indicated by participants during the study phase in Experiment 1a (**A**), Experiment 1b (**B**), Experiment 2a (**C**) and Experiment 2b (**D**). "h" and "m" respectively indicate 'highly' and 'moderately' with respect to either the expected or surprising congruency categories. Red dots represent means, and the middle horizontal bold lines indicate medians

Table 3 Item recognition accuracy analyses

Experiment	1a	1b	2a	2b
Fitting linear and quadratic model				
Linear comp. Fixed effects CI for β	**β =-.19; z=−4.1; *p* <.001 [-.28 : -.10]**	**β=-.14; z=−3.4; *p* <.001 [-.21 : -.05]**	**β=-.20; z=−4.0; *p* <.001 [-.30: -.10]**	**β=−0.21; z=−5.2; *p* <.001 [-.30 : -.13]**
Quadratic comp. Fixed effects CI for β	β=.04; z=.74; p=.46 [-.07 :.16]	β=.01; z=.1; p=.93 [-.11 :.12]	β=-.03; z=-.57; p=.57 [-.13 :.09]	**β=.17; z=3.5; *p* <.001 [.07 :.26]**
Particip. rand. Var	τ^2_0=.30; τ^2_1=.03;	τ^2_0=.54; τ^2_1=.02;	τ^2_0=.47; τ^2_1=.02;	τ^2_0=.21; τ^2_1=.03;
Item rand. var	τ^2_0=.70; τ^2_1=.02;	τ^2_0=.64; τ^2_1=.03;	τ^2_0=.71; τ^2_1=.10;	τ^2_0=.67; τ^2_1=.15;
Congruency range CI of minima point				[-.1.16 : 1.18] **[.21 :.1.07]**
DF	6320	6316	5144	8666
Old items				
Hits - means	.67 (.11)	.50 (.15)	.66 (.12)	.61 (.10)
Across categories	.73;.69;.65;.64	.55;.53;.49;.49	.71;.69;.65;.62	.68;.61;.56;.60
DK - means	.04 (.04)	.14 (.13)	.05 (.08)	.10 (.11)
Across categories	.02;.03;.05;.05	.13;.12;.15;.14	.05;.07;.05;.07	07;.09;.11;.10
Misses - means	.29 (.12)	.36 (.15)	.29 (.12)	.29 (.12)
Across categories	.25;.28;.30;.31	.32;.35;.36;.37	.24;.24;.30;.31	.25;.30;.33;.30
New items				
CR	.80 (.12)	.70 (.17)	.83 (.16)	.72 (.17)
False alarms	.06 (.08)	.06 (.07)	.06 (.08)	.05 (.08)
N participants	32	32	26	46

Results of logistic linear mixed effects model fit to accuracy in item recognition trials across experiments. β and z represent the coefficients of the linear and quadratic fixed components in the models; Particip. – participants; rand var. – random variable variances; comp – components; CI – confidence interval; $\tau^2_{0,}$ τ^2_1– respectively represent the intercept and slope variances of the random variables. In case of valid U-shaped we show the congruency continuous range and the CI range of the minima point; DF is the degrees of freedom. Across categories lines indicate the means values in each of the 4 congruency categories – from highly expected toward highly surprising. DK – don't know response; CR – correctly rejected responses; the numbers in parentheses show standard deviations (SD). N - Number of participants whose data was used in item recognition and recollection analyses, excluding those who performed at or above the chance level threshold

Item recollection and familiarity expressed in "remember" and "know" responses

Of those old items correctly recognised as old, the proportion receiving a "remember" (R) or "know" (K) judgment are shown in Fig. 3. Estimates of familiarity (F) are also shown, under the assumption that familiarity is independent of recollection (which is assumed to underlie "remember" responses; i.e., by expressing familiarity as the proportion of trials receiving "know" responses out of those studied trials that were not endorsed with a "remember" response; Yonelinas, 2002). In brief, recollection increased with greater expectedness, whereas familiarity mostly increased with greater surprise.

Trial-based logistic linear mixed effects models fit separately to recollection and familiarity across congruency categories showed a significant negative linear effect of recollection in every experiment, with an additional positive quadratic term in Experiment 2b. This quadratic effect does not indicate a U-shape pattern as the minima points are external to the congruency range. In addition, familiarity showed a significant positive linear effect in three of the four experiments. We report the fixed effects results and their confidence intervals, the random effects variances and the degrees of freedom. These results are shown in Table 4. The R-code model equation for recollection is shown below whereas the familiarity equation is similar. Additionally, we report in Table 4 the descriptive statistics of recollection and familiarity. R-code model equation for recollection:

$$model.recollect.sq = glmer(recollection \sim congruency + I(congruency^2)+ (1 + congruency|participant) + (1 + congruency|items), data.recollect, family = binomial)$$

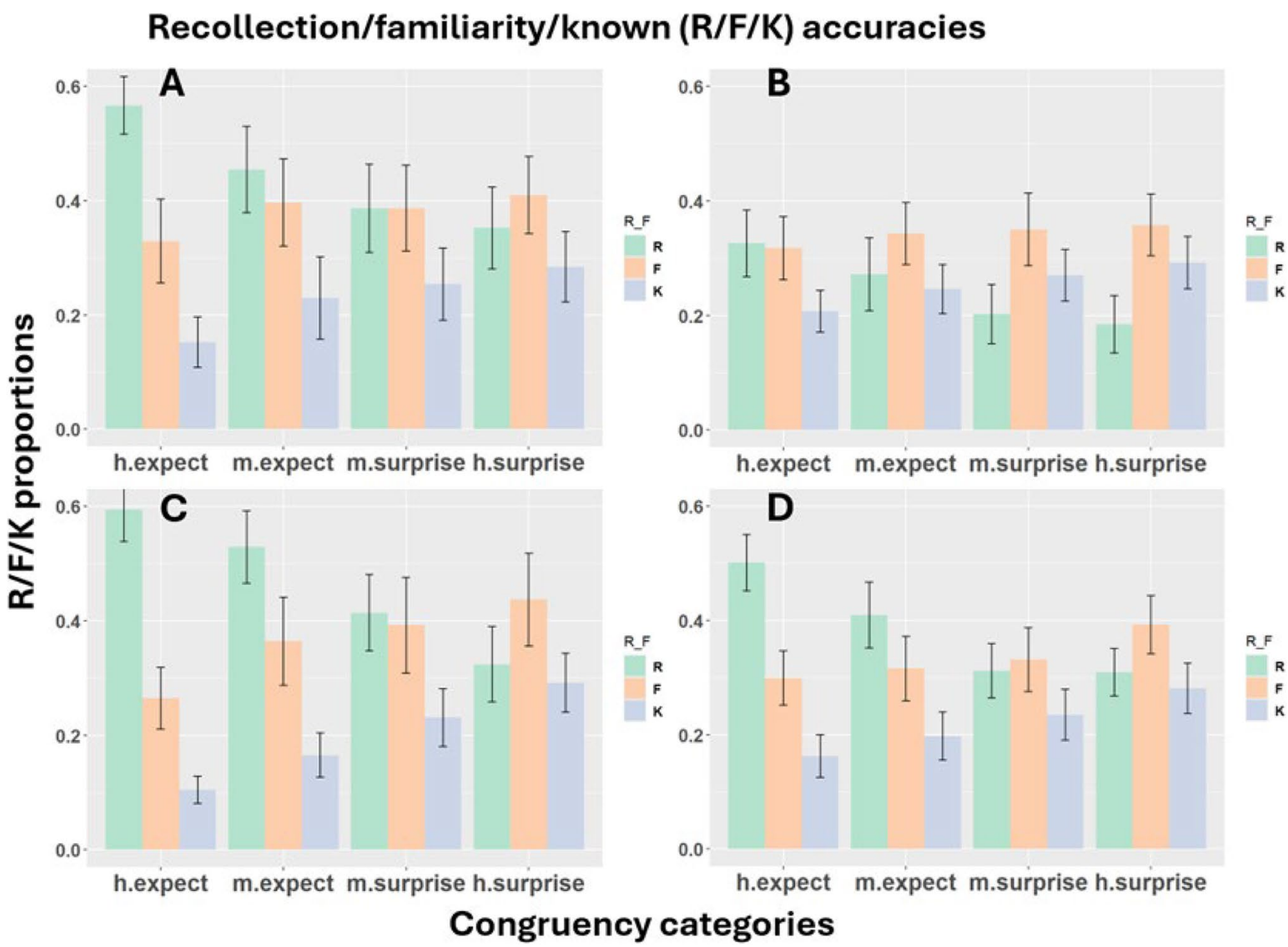

Fig. 3 Item Recognition Accurate Trials Mapped onto Recollection/Familiarity. *Notes*. The proportion of old items that were correctly recognized (hits) and subsequently judged to be remembered or known mapped onto recollection (R) and familiarity (F) measures and shown together with the known (K) proportions as a function of the congruency categories defined by each participant during the study phase for Experiment 1a (A), Experiment 1b (B), Experiment 2a (C) and Experiment 2b (D). "h" and "m" respectively indicate 'highly' and 'moderately' with respect to either the expected or surprising congruency categories. The error bars represent standard error of the mean

Source memory test results

Source memory for whether an object appeared in an expected or surprising context showed a U-shaped pattern in all four experiments (Fig. 4; i.e., high levels of expectedness or surprise were remembered better than moderate levels).

Trial-level logistic linear mixed effects models of source memory revealed a U-shape function across all experiments, driven by a significant quadratic component in all cases, with an internal minimum. The linear component did not show any trend. We report the fixed effects results and their confidence intervals, the random effects variances and the degrees of freedom (Table 5). Descriptive statistics are shown at the bottom of Table 5. The R-code model equation for source memory is shown below:

$$model.source.sq = glmer(sourceMemory \sim congruency + I(congruency^2) + (1 + congruency|participant) + (1 + congruency|items), data.source, family = binomial)$$

Table 4 Recollection/familiarity analyses

Experiment	1a	1b	2a	2b
Fitting model for recollection				
Linear comp. **CI for β**	**β=-.50; z=−6.9;** ***p* <.001** **[-.64 :-.35]**	**β=-.44; z=−5.9;** ***p* <.001** **[-.58 : -.29]**	**β=-.61; z=−9.7;** ***p* <.001** **[-.73 : -.49]**	**β=-.50; z=−9.6;** ***p* <.001** **[-.60 : -.40]**
Quadratic comp. **CI for β**	β=.06; z=1.4; *p* =.17 [-.03 :.19]	β=.01; z=.10; *p* =.95 [-.13 :.13]	β=-.05; z=-.78; *p* =.43 [-.16 :.07]	β=.10; z=1.87; *p* =.06 [-.0:.19]
Particip. rand. var **Item rand. var**	τ^2_0=.68; τ^2_1=.13; τ^2_0=53; τ^2_1=.07;	τ^2_0=.85; τ^2_1=.12; τ^2_0=.68; τ^2_1<.01;	τ^2_0=.36; τ^2_1=.06; τ^2_0=.71; τ^2_1=.10;	τ^2_0=.50; τ^2_1=.07; τ^2_0=.70; τ^2_1=.15;
DF	6320	6316	5144	8666
Means across categories	.57;.46;.39;.35	.33;.27;.20;.18	.59;.53;.41;.32	.52;.41;.31;.31
Fitting model for familiarity				
Linear comp. CI for β	**β=.15; z=2.5;** ***p* <.05** **[.03 :.27]**	β=.08; z=1.6; *p* =.11 [-.02 :.17]	**β=.33; z=4.2;** ***p* <..001** **[.17 :.48]**	**β=.18; z=3.3;** ***p* <..01** **[.07 :.29]**
Quadratic comp. **CI for β**	β=-.07; z=-.8; *p* =.41 [-.21:.08]	β=-.06; z=-.86; *p* =.39 [-.18 :.07]	β=-.12; z=−1.6; *p* =.12 [-.27 :.03]	β=.08; z=1.2; *p* =.22 [-.05 :.20]
Parτicip rand. var **Item rand. var**	τ^2_0=.66; τ^2_1=.05; τ^2_0=.45; T^2_1<.01;	τ^2_0=.44; τ^2_1=.03; τ^2_0=.47; τ^2_1=.02;	τ^2_0=.30; τ^2_1=.09; τ^2_0=.35; τ^2_1=.03;	τ^2_0=.45; τ^2_1=.07; τ^2_0=.33; τ^2_1=.05;
DF	3504	4742	2830	5248
Across categories	.33;.40;.39;.41	.32;.34;.35;.36	.26;.36;.39;.44	.32;.33;.34;.41

Results from trial level logistic linear mixed effects models fit separately to recollection and familiarity across experiments. Recollection declined monotonously from expectedness to surprise across all experiments whereas familiarity showed the opposite pattern of increase across this range in three of the four experiments. β and z represent the coefficients of the linear and quadratic components in the models; Particip. – participants; rand var. – random variable variances; comp – components; CI – confidence interval; $\tau^2_{0,}$ τ^2_1– respectively represent the intercept and slope variances of the random variables. DF is the degrees of freedom; Across categories lines indicate the means values in each of the 4 congruency categories – from highly expected toward highly surprising

To assess the possibility that some accurate responses reflected guessing, we estimated the guess proportions within correct responses using multinomial processing trees (MPT), similar to the item recognition analysis (without new items). Optimized estimates indicated low percentages of guesses within accurate responses (<6%) (see Supplemental Material Table S2, Figs. S5–S6). We further tested the robustness of these estimates by randomly removing 6% of correct trials and repeating the linear plus quadratic statistical tests (n = 100 iterations). The quadratic component remained significant across all four experiments, confirming the robustness of the U-shaped pattern to guesses (see Supplemental Material Fig. S7).

To test whether the item recognition test influenced the subsequent source memory test, we fitted generalized linear mixed-effects models (GLMMs) predicting source memory accuracy from item recognition accuracy. The additional variance explained (differences of R^2) by item accuracy was 0.09 in Experiment 1a, 0.13 in Experiment 1b, 0.10 in Experiment 2a and 0.15 in Experiment 2b. The model specification, analyses and full results are shown in Table S3 in the Supplemental Material. Next, we repeated the source memory analysis while limiting the analyzed trials to those in which participants provided accurate responses to old items in the recognition test. Our results still confirmed the U-shaped pattern in all four experiments, with very similar results (see the results in Table S4 in the Supplemental Material).

An interesting further observation in Fig. 4 is the wider spread of source memory accuracy across participants for expected relative to surprising events. This difference in variance was confirmed statistically by Levene's test across all four experiments (Table 5). Moreover, performance in the congruent conditions appeared to be bimodal, with one subgroup performing at levels comparable to the surprising conditions and others performing lower (see distribution of gray lines in "h.expect" condition).

Post hoc power analysis

Following reviewer concerns regarding the appropriateness of G*Power for our experimental designs, we conducted a simulation-based power analysis. G*Power assumes a single observation per participant per condition, and its default repeated-measures ANOVA implementation further inflates sample size estimates by assuming a "double dissociation"

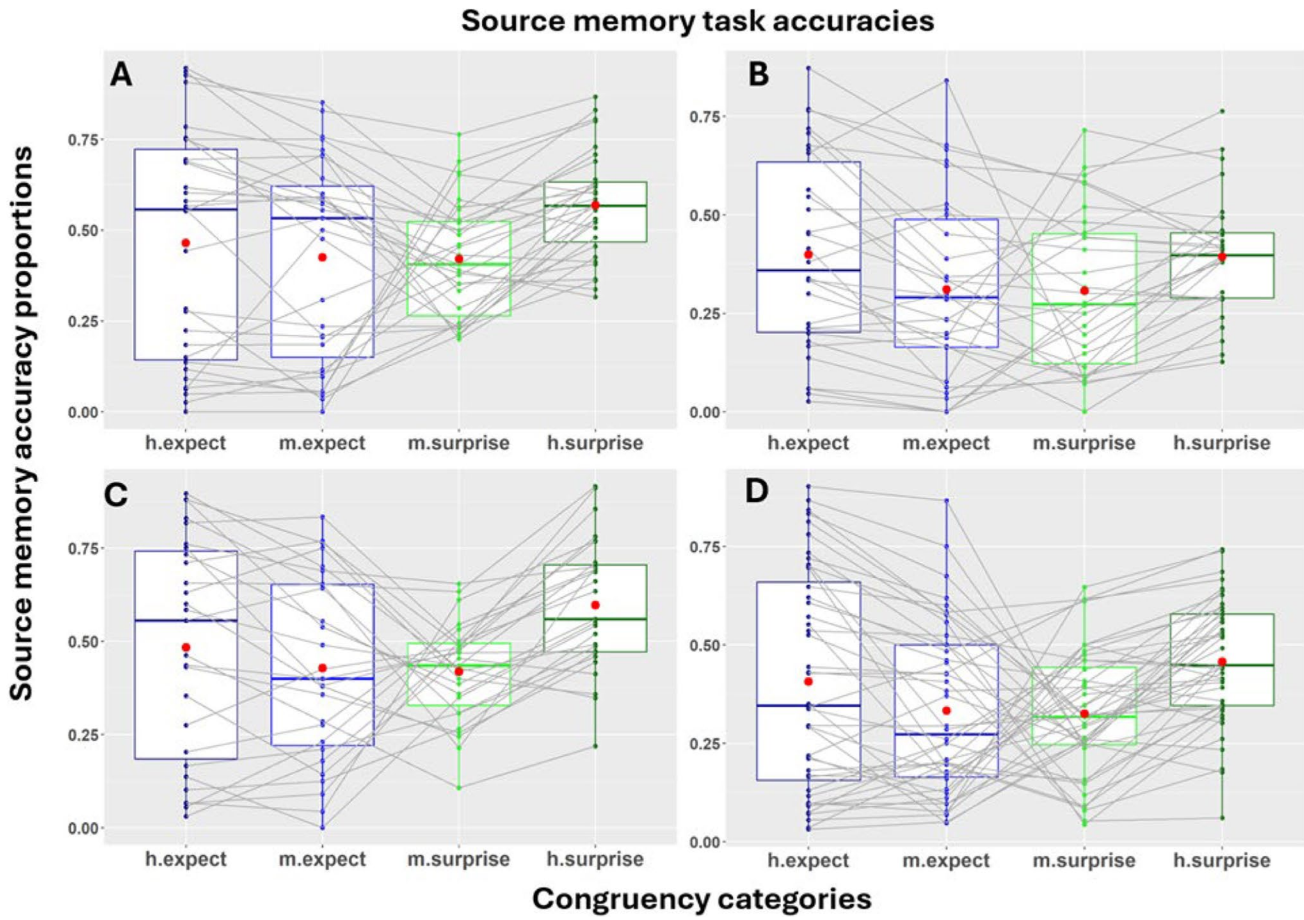

Fig. 4 Source Memory Accuracies. *Notes*. Proportion of source memory accuracy judgments as function of the four congruency categories that were endorsed by participants during the study phase in Experiment 1a (A), Experiment 1b (B), Experiment 2a (C) and Experiment 2b (D). "h" and "m" respectively indicate 'highly' and 'moderately' with respect to either the expected or surprising congruency categories. Red dots represent means, and the middle horizontal bold lines indicate medians

effect structure (Thibault et al., 2024). Neither assumption holds for the present designs, in which participants contributed many trials per condition (Brysbaert & Stevens, 2018).

To avoid the circularity of estimating power from our own observed effects, we based the simulations on effect sizes drawn from an independent study. We used Ortiz-Tudela et al. (2024), whose paradigm and mixed-model analysis of source and item memory closely parallel our own, and whose young-adult ending-recognition and cue-recognition tasks provide suitable independent estimates of the quadratic congruency effect and the linear item-recognition effect, respectively. From their reported polynomial components, we derived fixed-effect coefficients on the log-odds scale of our own predictor and substituted these into each experiment's fitted model, retaining its design and the estimated random-effects structure. Power was estimated using a simulation procedure equivalent to that of the *simr* package in R (Green & MacLeod, 2016; Kumle et al., 2021) implemented directly in R to accommodate independently generated fixed effects. Simulations of power were based on the fitted *glmer* model for source memory accuracy, focusing on the quadratic congruency effect in source memory—the primary fixed effect of interest—and on the linear fixed effect in item recognition. Simulations were run across a range of participants sample sizes, with 100 iterations per sample size point. See detailed procedure in the Supplemental Material.

Across all four experiments, 80% power was reached at sample sizes below those we planned and collected: for both the quadratic congruency effect in source memory ($N =$ 7–15) and the linear effect in item recognition ($N =$ 14–22), confirming that the designs were adequately powered to detect effects of the magnitude reported by Ortiz-Tudela et al. (2024) (see Table S10 and Fig. S10 in the Supplemental Material).

Table 5 Source memory accuracy analyses

Experiment	1a	1b	2a	2b
Fitting linear and quadratic model				
Linear comp. **CI for β**	β=.42; z=1.2; *p* =.21 [-.14 :.62]	β=-.01; z=.07; *p* =.94 [-.33 :.35]	β=.44; z=1.67; *p* =.10 [-.08 :.95]	β=.10; z=.45; *p* =.66 [-.34 :.53]
Quadratic comp. **CI for β**	**β=.30; z=5.7;** ***p* <.001** **[.20 :.41]**	**β=.30; z=5.5;** ***p* <.001** **[.19 :.41]**	**β=.39; z=7.4;** ***p* <. 001** **[.29 :.50]**	**β=.37; z=8.0;** ***p* <. 001** **[.28 :.46]**
Congruency range **CI of minima point**	[−1.17 : 1.18] **[−1.04 :.24]**	[−1.13 : 1.19] **[-.58 :.54]**	[−1.16 : 1.18] **[.56 : 1.78]**	[−1.15 : 1.18] **[-.72 :.46]**
Particip. rand. var **Item rand. var**	τ^2_0=.60; τ^2_1=.35; τ^2_0=.17; τ^2_1=.20;	τ^2_0=.56; τ^2_1=.13; τ^2_0=.15; τ^2_1=.44;	τ^2_0=.41; τ^2_1=.06; τ^2_0<.01; τ^2_1<.01;	τ^2_0=.22; τ^2_1=.13; τ^2_0=.26; τ^2_1=.36;
DF	6320	6117	5248	8468
Levene's test: **Compared** **Variances**	**F=30.4;** ***p* <.001;** DF=62	**F=6.0;** ***p* <.05;** DF=60	**F=11.9;** ***p* <.01;** DF=52	**F=24.8;** ***p* <.001;** DF=84
Descriptive statistics				
Hits – means **Across categories**	.49 (.17) .46;.42;.43;.57	.37 (.16) .40;.32;.30;.38	.50 (.15) .48;.43;.42;.60	.40 (.14) .41;.32;.34;.46
DK – means **Across categories**	.32 (.21) .39;.34;.31;.26	.46 (.21) .45;.47;.43;.41	.33 (.12) .39;.40;.27;.27	.44 (.18) .48;.49;.44;.39
Missed – means **Across categories**	.19 (.09) .15;.24;.26;.17	.17 (.07) .15;.21;.28;.21	.17 (.07) .13;.17;.31;.13	.16 (.08) .11;.19;.22;.15
N participants	31	30	27	46

Results of logistic linear mixed effects models fit to source memory accuracy measures across congruency categories showing U-shaped patterns in all four experiments. β and z represent the coefficients of the linear and quadratic components in the models; Particip. – participants; rand var. – random variable variances; comp. – components; CI – confidence interval; $\tau^2_{0,}$ τ^2_1– respectively represent the intercept and slope variances of the random variables. In case of valid U-shaped we show the congruency continuous range and the CI range of the minima point; DF is the degrees of freedom. The variances of congruent versus incongruent conditions (collapsed across high/moderate) were compared using Levene's Test. DK – don't know; Across categories lines indicate the means values in each of the 4 congruency categories – from highly expected toward highly surprising

Discussion

We examined the SLIMM framework's prediction that both schema-violating (surprising) and schema-compliant (expected) experiences provide mnemonic benefits compared with moderately surprising or moderately expected experiences, and that surprising events should benefit from greater recollection of their contextual details (source memory). Participants viewed scene images containing either congruent or incongruent objects and categorized their level of surprise or expectedness. Following this encoding, memory was assessed through an item recognition (old/new) test, further categorized by recollection and familiarity; and a source memory test, in which participants recalled their expectedness/surprise selection at encoding. Contrary to SLIMM, item recognition showed only an advantage for expected objects, which was driven by recollection (not familiarity), while source memory results showed a U-shaped function across congruency levels (not an advantage for surprising trials only). In Experiments 2a and 2b, we also administered an associative memory test, in which participants identified the scene in which an object had been embedded, but interpretation of these results is hindered by a confound; when we controlled for this confound in Experiment S1, an advantage emerged for surprising trials, though this pattern was presumably biased by foil-target competition and depended on overall performance level (see Supplementary Material).

Advantage for expected objects in the item-recognition test

Across Experiments 1a, 1b, 2a, and 2b, item recognition was higher for expected than for surprising trials. Modelling the proportion of guesses within accurate responses revealed a negligible percentage of guesses in all four experiments, suggesting this is not simply a guessing bias. These results replicate findings from several previous studies (Alejandro et al., 2021; Bein et al., 2015; Frank et al., 2018; Packard et al., 2017; Packard & Soto-Faraco, 2025; Staresina et al., 2009; van Kesteren et al., 2013a; though see Ortiz-Tudela et al., 2024), which reported an advantage for recognizing items previously encountered in a congruent context, which in our case is the background scene, compared with an incongruent context. This pattern was evident in both immediate and delayed (next-day) tests.

Moreover, this memory advantage was linked to recollection. One explanation for this recollection-based congruency advantage is that participants generate semantic associations of the object probes presented at test (Barsalou, 1982; Kutas & Federmeier, 2000; Tulving & Thomson, 1973). These associations are more likely to facilitate recollection of a previously studied congruent context than an incongruent one (Bein et al., 2015; Craik & Tulving, 1975). This recollection advantage could arise from two sources: First, the item probe may cue a semantic associate that appeared in the scene, thereby triggering retrieval of that scene (henceforth, "scene source"). Second, the item probe may cue a semantic associate that was not present in the scene but was nonetheless generated during retrieval ("schema-related source"). Both types of retrieval are likely to be experienced as recollection (see also Craik & Tulving, 1975; Moscovitch & Craik, 1976; Steyvers & Hemmer, 2012). In contrast, generating semantic associations to a cue is less likely to facilitate recollection during an unexpected trial, because such associations are less likely to have been formed during study within an incongruent scene. One limitation concerns the remember/know procedure: because participants judged only objects they had correctly identified as old, the indirect accuracy feedback may have inflated "remember" relative to "know" responses. We cannot rule out this bias, but as it would apply across all congruency levels, it is unlikely to account for the recollection differences between expected and surprising trials.

This congruency bias may have obscured potential benefits for the encoding of highly surprising objects, which could arise from prediction error or heightened attention at study (Brewer & Treyens, 1981; Greve et al., 2017; Quent et al., 2021). The left-tilted U-shape in Experiment 2b may indicate that such obscuring was sometimes partial. Moreover, the strategy of generating schema-related (or semantically related) associations at retrieval does not always facilitate memory and is not always feasible. For example, some previous studies found the opposite pattern (to our data) of higher recollection of unexpected trials (Cycowicz et al., 2008; Kafkas & Montaldi, 2018; W.-C. Wang et al., 2018). We suspect that the reason for the discrepancy is that the use of semantically-generated cues at retrieval did not help performance in those studies. In Kafkas and Montaldi (2018) for example, the critical test involved remembering the order of object presentations, a task for which schema-related retrieval offers little benefit (see also Greve et al., 2017; Kim et al., 2014; Sinclair & Barense, 2018). In Cycowicz et al. (2008), objects were studied with a consistent outline-color (e.g., a yellow-outlined banana) or an inconsistent outline-color (e.g., a blue-outlined banana). At test they were all probed with a black neutral outline. Item recollection was higher for inconsistent (surprising) probes, likely because the canonical schema color of objects is unambiguous (e.g., yellow for banana) such that schema-related retrieval could not provide any further schema associations to benefit recollection for consistent (expected) trials. Likewise, W.-C. Wang et al. (2018) probed memory using a complete stimulus, rendering schema-related retrieval ineffective. Other demonstrations of memory benefits for schema-inconsistent over schema-consistent objects in item-recognition may have similar explanations. For example, after 45 s of studying an office environment, participants recognized inconsistent objects better than consistent ones (Mäntylä & Bäckman, 1992). Because all items were studied within the same physical context (an office), schema activation at retrieval would apply broadly rather than selectively facilitating individual congruent objects (Moscovitch & Craik, 1976; Steyvers & Hemmer, 2012). Therefore, the distinctiveness of schema-inconsistent objects at encoding likely explained their better recognition. In other situations, distinctiveness or surprise-driven encoding may be strong enough to overcome any schema-related retrieval strategy. For example, Erdfelder and Bredenkamp (1998) used sentences, in which schema-inconsistent sentences were two or three times less frequent than schema-consistent ones. This rarity may have increased attention, outweighing any schema-related retrieval advantage for consistent ones.

Assuming familiarity is independent of recollection (Diana et al., 2006; Yonelinas, 2002), we found higher levels of familiarity for surprising than expected objects in three of our four experiments. This is different from other experiments in which familiarity showed no reliable congruency effect (Quent et al., 2022; Raykov et al., 2026). The reason for this difference is unclear.

Source memory benefits from highly surprising and highly expected events

Across all four experiments, we consistently observed a U-shaped function in the source memory test, in which participants retrieved whether an object had been expected or was surprising at study: Memory performance was highest for both highly surprising and highly expected trials compared to moderately surprising or expected trials. However, we also observed greater variability in accuracy across participants for expected trials compared with surprising ones: While some participants accurately remembered that an object had been paired with a congruent scene, others performed poorly. This variability suggests that participants may have differed in their retrieval strategies (Unsworth, 2017), perhaps reflecting distinct retrieval orientations (Herron & Wilding, 2004; Rugg & Wilding, 2000). Again, we do not think this result reflects guessing, because the percentage of such guesses was estimated as <6% in all experiments. We also assessed whether completing the item recognition test influenced subsequent source memory performance and found minimal impact: (a) correlating the variance estimates from the statistical models between the two tests yielded $R^2 < 0.16$, below the 0.25 threshold for weak shared variance, indicating independent performance across tests (Berchtold, 2016; Koo & Li, 2016; Matheson, 2019); (b) the U-shaped patterns in source memory were sustained when analyses were limited to trials with accurate item recognition responses.

The proposed "tag" account of source memory

Participants were instructed to retrieve their feeling of expectedness or surprise at encoding, rather than remembering the specific scene paired with each object. The absence of an item recognition advantage for surprising items, combined with overall higher performance in item recognition compared to source memory, suggests that memory performance for congruent trials was specifically reduced in the source memory test. In other words, participants often failed to remember that an object had appeared in an expected context during study, even when they could recognize that object in the item recognition test.

One possible explanation is that participants attach a tag for "expected" or "surprising" to encoding trials, and that recalling an "expected" tag is more difficult than recalling a "surprising" tag. Note that when recalling these event tags (i.e., expectedness or surprisingness), people do not necessarily retrieve other episodic details, such as the scene itself. Previous research on eyewitness memory suggests that individuals may recall an event as unexpected but fail to remember its details (Tuckey & Brewer, 2003). Similarly, research on hindsight bias shows that initial surprise can be remembered while the reasons for that surprise are lost through mental sense-making (Pezzo, 2003).

These tags would be more likely to be created at study for highly surprising and for highly expected events than for moderate ones, producing the observed U-shaped memory pattern. Information at the extremes (perfect schema match or complete surprise) tends to be clearer and more distinctive than moderate information, and is therefore better encoded (Gretz & Huff, 2020; Icht et al., 2014). Our study RTs fit with this view, showing an inverted U-shaped pattern across congruency, with faster RTs for highly expected and highly surprising items and slowest RTs for the middle items (see also Kiran & Thompson, 2003; T. Wang et al., 2016). Furthermore, highly surprising items were slower than highly expected items at study, reflecting an incongruency cost commonly observed during encoding (Head & Pedoe, 1990; Yacoby et al., 2023), likely due to large prediction errors and extra attention that benefits encoding (Brewer & Treyens, 1981; Küppers & Bayen, 2014). Consequently, these "surprisingness" tags may be more salient than the "expectedness" tags when acting as retrieval cues (Graesser et al., 1979; Graesser & Nakamura, 1982; McDaniel et al., 1995).

The assumed lower salience of expectedness tags may explain the larger variance observed among participants in this category. Participants may have employed different retrieval strategies when recalling the congruency of expected events: Although participants were instructed to recall only the degree of expectedness, it is unknown on what they based their judgments. Some may have relied directly on expectedness tags when these were sufficiently salient, while others may have attempted the schema-related strategy described earlier. This schema-related strategy may have produced poorer source memory performance than direct tag retrieval, since tag-based retrieval requires accessing a single piece of binary information, whereas the semantic strategy requires a cascade of successful and relevant associations related to the scene at study, making the probability of success lower (Tversky & Kahneman, 1983). For example, consider a fishing rod presented in a riverbank scene. Participants could have solved this either by recalling the expectedness tag directly, or by retrieving associations with various congruent scenes—lake, river, hobby shop—and attempting to match the fishing rod to each context until the original study scene was recalled and then determining whether the item was expected or surprising. Mixing the poorer performance of this multi-step semantic approach with the simpler, higher-performance tag-based approach would result in the larger variance observed for expected versus surprising categories in source memory retrieval. Future studies could further examine this increased variance in light of this, and alternative, explanations.

The formation of these tags may be particularly likely in our study because the encoding task-oriented participants toward evaluating expectancy and surprise. Indeed, repeating similar cognitive processes at both study and test (in our case, evaluating expectedness and surprise) generally facilitates memory (Guo et al., 2021; Jacoby et al., 2005; Salan et al., 2025; Schulman, 1974; Zhang & Hupbach, 2023). Support for this explanation comes from Ortiz-Tudela et al. (2024), who found a similar U-shaped pattern in memory accuracy when at study participants were oriented to the expectedness of a cartoon ending, and at test were asked to recognize which of three possible sequence endings had been presented at study: expected, surprising, or neutral. Memory was better for expected and surprising endings compared to neutral ones, consistent with the suggested tag-based retrieval account.

Comparing other source memory evidence to our findings

Several previous studies have reported schema inconsistency effects (Ehrenberg & Klauer, 2005), in which source memory showed an advantage for unexpected versus expected contexts. In two studies, for example, participants were asked to remember whether an object was studied in an expected or unexpected context, or whether it was new (Küppers & Bayen, 2014; Schaper et al., 2019). The inconsistency effects in these studies were attributed to the attentional elaboration hypothesis (see below in the alternative accounts Discussion subsection). Since these studies compared only extreme expected and unexpected contexts, without intermediate levels, detection of a U-shaped pattern was not possible. However, their finding that source memory for unexpected events was better than for expected events does not contradict our suggested tag account, which explains memory benefits at the extremes of the congruency continuum relative to moderate values. The specific direction of the effect between congruency extremes is not central to this account and may depend on additional factors, such as distinctiveness.

To what extent do our data align with SLIMM and alternative accounts?

While the SLIMM framework guided the present hypotheses and experiments, it is clearly not sufficient to explain the full pattern of results: the recollection advantage of expected over surprising trials in item recognition contradicts SLIMM's predictions, and the U-shaped pattern we observed in source memory is not aligned with SLIMM's prediction of memory benefits solely in surprising trials (van Kesteren et al., 2012). Nonetheless, if we allow for the suggested schema-related retrieval strategy and for tag encoding and retrieval processes, it becomes possible to account for most findings across memory tests and experiments. This framework may also account for findings in previous studies, such as the recollection advantage for expected trials reported by Quent et al. (2022) and the U-shaped source memory pattern observed in Ortiz-Tudela et al. (2024) (see above). A key contribution of our experiments is the demonstration that memory is a complex function of both expectedness/surprise and the type of memory test. By contrast, our proposed framework requires further empirical research to evaluate and refine its predictions.

Alternative accounts predict a memory advantage for schema-inconsistent over schema-consistent episodes that originate during encoding. For example, the attentional elaboration hypothesis (Brewer & Treyens, 1981; Friedman, 1979; Loftus & Mackworth, 1978) attribute this benefit to increased attention triggered by inconsistent information (McDaniel et al., 1995), which outweighs potential memory benefits from the easier assimilation and shallower encoding of schema-consistent information. Importantly, this account (and see also the copy-and-tag account in Introduction) focused exclusively on comparing the extremes of highly inconsistent and highly consistent events, without addressing intermediate congruency levels, as SLIMM does, nor does it address differences between retrieval test effects and retrieval strategies as proposed in our framework. Consequently, the schema-consistent memory benefits observed in the item recognition test contradict these accounts' predictions, as they did not consider the semantic retrieval strategies that may facilitate retrieval of schema-consistent information. Likewise, our U-shaped source memory findings (or those reported in Ortiz-Tudela et al., 2024) are similarly left unexplained.

Conclusion

In this study, we investigated the effects of schema-congruent and schema-violating experiences on memory, guided by the SLIMM framework. Our data showed that participants' perception of trials as highly or moderately expected or surprising, affected memory performance in a complex manner that depended on the type of memory test. Although we cannot uniquely attribute effects to retrieval strategies, we hypothesize that performance depended on two factors: first, whether participants directly remembered the object itself in item recognition or instead relied on a schema-related retrieval strategy to reconstruct the studied scene (the latter would disproportionately benefit expected objects and produce a recollection-based memory advantage), and second, whether in source memory, participants encoded and retrieved event tags marking whether the scene context had been expected or surprising. In the absence of such tags, they would fall back on the schema-related scene reconstruction, together accounting for the U-shaped source memory pattern. Thus, theoretical frameworks such as SLIMM require further elaboration to fully account for the influence of prior knowledge across different memory tasks and the retrieval strategies they may engage.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-026-01930-1.

Author contributions A.Y.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing—original draft. A.G.: Conceptualization, Data curation, Methodology, Supervision, Visualization, Writing—review & editing. E.S.: Data curation, Investigation, Methodology, Project administration. L.S.: Project administration, Resources, Software, Supervision, Visualization. R.H.: Conceptualization, Investigation, Methodology, Software, Visualization, Writing—review & editing.

D.A.L.: Conceptualization, Funding acquisition, Methodology, Supervision, Visualization, Writing—review & editing.

Funding Open access funding provided by Reichman University. Andrea Greve and Richard Henson were supported by UKRI Medical Research Council (MRC) [Grant No. SUAG/086/G116768].

Data availability All study materials and collected data are publicly available on our OSF site (https://osf.io/u8fzy/files/osfstorage). The study was not preregistered.

Code availability Our analyses' code is publicly available on our OSF site (https://osf.io/u8fzy/files/osfstorage).

Declarations

Ethics approval and consent to participate Are elaborated in the Participant's subsection of the Methods section.

Consent for publication For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

Conflict of interest We have no known conflict of interest to disclose.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203

Alejandro, R. J., Packard, P. A., Steiger, T. K., Fuentemilla, L., & Bunzeck, N. (2021). Semantic congruence drives long-term memory and similarly affects neural retrieval dynamics in young and older adults. *Frontiers in Aging Neuroscience, 13*, Article Article 683908. https://doi.org/10.3389/fnagi.2021.683908

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10*(1), 82–93. https://doi.org/10.3758/BF03197629

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

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

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

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the effect of semantic congruency on episodic memory: The role of integration and relatedness. *PLoS ONE, 10*(2), Article Article e0115624. https://doi.org/10.1371/journal.pone.0115624

Bein, O., Plotkin, N. A., & Davachi, L. (2021). Mnemonic prediction errors promote detailed memories. *Learning & Memory, 28*(11), 422–434. https://doi.org/10.1101/lm.053410.121

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations*. https://doi.org/10.1177/2059799116672875

Brainerd, C. J., Chang, M., Liu, X., & Bialer, D. M. (2025). Memory framing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 51*(12), 1941–1952. https://doi.org/10.1037/xlm0001485

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*(2), 207–230. https://doi.org/10.1016/0010-0285(81)90008-6

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*. https://doi.org/10.5334/joc.10

Chua, E. F., Hannula, D. E., & Ranganath, C. (2012). Distinguishing highly confident accurate and inaccurate memory: Insights about relevant and irrelevant influences on memory confidence. *Memory, 20*(1), 48–62. https://doi.org/10.1080/09658211.2011.633919

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Cycowicz, Y. M., Nessler, D., Horton, C., & Friedman, D. (2008). Retrieving object color: The influence of color congruity and test format. *NeuroReport, 19*(14), 1387–1390. https://doi.org/10.1097/WNR.0b013e32830c8df1

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13*(1), 1–21. https://doi.org/10.3758/bf03193807

Dobbins, I. G., & McCarthy, D. (2008). Cue-framing effects in source remembering: A memory misattribution model. *Memory & Cognition, 36*(1), 104–118. https://doi.org/10.3758/MC.36.1.104

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41*(4), 369–387. https://doi.org/10.1016/j.jesp.2004.08.001

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

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

Erdfelder, E., & Bredenkamp, J. (1998). Recognition of script-typical versus script-atypical information: Effects of cognitive elaboration. *Memory & Cognition, 26*(5), 922–938. https://doi.org/10.3758/BF03201173

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

Frank, D., Montaldi, D., Wittmann, B., & Talmi, D. (2018). Beneficial and detrimental effects of schema incongruence on memory for contextual events. *Learning & Memory, 25*(8), 352–360. https://doi.org/10.1101/lm.047738.118

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316–355. https://doi.org/10.1037/0096-3445.108.3.316

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229–244). Oxford University Press. https://doi.org/10.1093/oso/9780195122657.003.0015

Graesser, A. C., Gordon, S. E., & Sawyer, J. D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *Journal of Verbal Learning and Verbal Behavior, 18*(3), 319–332. https://doi.org/10.1016/S0022-5371(79)90182-8

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. *Psychology of Learning and Motivation, 16*, 59–109. https://doi.org/10.1016/S0079-7421(08)60547-2

Gray, S. J., & Gallo, D. A. (2015). Disregarding familiarity during recollection attempts: Content-specific recapitulation as a retrieval orientation strategy. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 41*(1), 134–147. https://doi.org/10.1037/a0038363

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Gretz, M. R., & Huff, M. J. (2020). Multiple species of distinctiveness in memory? Comparing encoding versus statistical distinctiveness on recognition. *Memory, 28*(8), 984–997. https://doi.org/10.1080/09658211.2020.1803916

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language, 94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology. General, 148*(2), 325–341. https://doi.org/10.1037/xge0000498

Guo, J., Shubeck, K., & Hu, X. (2021). Relationship between item and source memory: Explanation of connection-strength model. *Frontiers in Psychology, 12*, Article 691577. https://doi.org/10.3389/fpsyg.2021.691577

Head, A. S., & Pedoe, D. S. T. (1990). A choice-reaction time task utilizing congruent and incongruent colour/word stimuli: Another reaction to Stroop. *Perceptual and Motor Skills, 71*(1), 115–119. https://doi.org/10.2466/pms.1990.71.1.115

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus, 20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Herron, J. E., & Wilding, E. L. (2004). An electrophysiological dissociation of retrieval mode and retrieval orientation. *NeuroImage, 22*(4), 1554–1562. https://doi.org/10.1016/j.neuroimage.2004.04.011

Hunt, R. R., & Worthen, J. B. (Eds.). (2006). *Distinctiveness and memory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.001.0001

Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.00886

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*(5), 852–857. https://doi.org/10.3758/BF03196776

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

Kafkas, A., & Montaldi, D. (2018). Expectation affects learning and modulates memory experience at retrieval. *Cognition, 180*, 123–134. https://doi.org/10.1016/j.cognition.2018.07.010

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences of the United States of America, 111*(24), 8997–9002. https://doi.org/10.1073/pnas.1319438111

Kiran, S., & Thompson, C. K. (2003). Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language, 85*(3), 441–450. https://doi.org/10.1016/S0093-934X(03)00064-6

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuhlmann, B. G., & Touron, D. R. (2012). Mediator-based encoding strategies in source monitoring in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1352–1364. https://doi.org/10.1037/a0027863

Kumle, L., Võ, M. L., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology, 67*(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

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

Liu, Z.-X., Grady, C., & Moscovitch, M. (2018). The effect of prior knowledge on post-encoding brain connectivity and its relation to subsequent memory. *NeuroImage, 167*, 211–223. https://doi.org/10.1016/j.neuroimage.2017.11.032

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 565–572. https://doi.org/10.1037/0096-1523.4.4.565

Lorini, E., & Castelfranchi, C. (2006). The unexpected aspects of surprise. *International Journal of Pattern Recognition and Artificial Intelligence, 20*(06), 817–833. https://doi.org/10.1142/S0218001406004983

Mäntylä, T., & Bäckman, L. (1992). Aging and memory for expected and unexpected objects in real-world settings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(6), 1298–1309. https://doi.org/10.1037/0278-7393.18.6.1298

Maril, A., Avital, R., Reggev, N., Zuckerman, M., Sadeh, T., Sira, L. B., & Livneh, N. (2011). Event congruency and episodic encoding: A developmental fMRI study. *Neuropsychologia, 49*(11), 3036–3045. https://doi.org/10.1016/j.neuropsychologia.2011.07.004

Matheson, G. J. (2019). We need to talk about reliability: Making better use of test-retest studies for study design and interpretation. *PeerJ, 7*, Article e6918. https://doi.org/10.7717/peerj.6918

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 422–435. https://doi.org/10.1037/0278-7393.21.2.422

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition, 21*(3), 1435–1455. https://doi.org/10.1016/j.concog.2012.04.014

Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15*(4), 447–458. https://doi.org/10.1016/S0022-5371(76)90040-2

Ortiz-Tudela, J., Nolden, S., Pupillo, F., Ehrlich, I., Schommartz, I., Turan, G., & Shing, Y. L. (2023). Not what u expect: Effects of prediction errors on item memory. *Journal of Experimental Psychology: General, 152*(8), 2160–2176. https://doi.org/10.1037/xge0001367

Ortiz-Tudela, J., Turan, G., Vilas, M., Melloni, L., & Shing, Y. L. (2024). Schema-driven prediction effects on episodic memory across the lifespan. *Philosophical Transactions of the Royal Society B: Biological Sciences, 379*(1913), Article 20230401. https://doi.org/10.1098/rstb.2023.0401

Packard, P. A., Rodríguez-Fornells, A., Bunzeck, N., Nicolás, B., de Diego-Balaguer, R., & Fuentemilla, L. (2017). Semantic congruence accelerates the onset of the neural signals of successful memory encoding. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 37*(2), 291–301. https://doi.org/10.1523/JNEUROSCI.1622-16.2016

Packard, P. A., & Soto-Faraco, S. (2025). Crossmodal semantic congruence and rarity improve episodic memory. *Memory & Cognition, 53*(5), 1396–1418. https://doi.org/10.3758/s13421-024-01659-9

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pezzo, M. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory, 11*(4/5), 421–441. https://doi.org/10.1080/09658210244000603

Quent, J. A., Greve, A., & Henson, R. N. (2022). Shape of U: The nonmonotonic relationship between object-location memory and expectedness. *Psychological Science, 33*(12), 2084–2097. https://doi.org/10.1177/09567976221109134

Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory, 179*, Article 107382. https://doi.org/10.1016/j.nlm.2021.107382

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org

Raykov, P. P., Gurunandan, K., Moniri, M., Jain, A., Greve, A., & Henson, R. N. (2026). Predictions and declarative memory encoding: Two fMRI paradigms provide slim pickings for SLIMM. *Philosophical Transactions of the Royal Society B: Biological Sciences, 381*(1954), Article 20250250. https://doi.org/10.1098/rstb.2025.0250

Roediger, H. L., Tekin, E., & Lin, W. (2022). Applying confidence-accuracy characteristic plots to recognition memory. In T. F. Brady & W. A. Bainbridge (Eds.), *Visual memory* (pp. 335–349). Routledge. https://doi.org/10.4324/9781003158134-20

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4*(3), 108–115. https://doi.org/10.1016/S1364-6613(00)01445-5

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Routledge.

Salan, J., Smith, D. E., Shafer, E. S., & Diana, R. A. (2025). Variation in encoding context benefits item recognition. *Memory & Cognition, 53*, 725–745. https://doi.org/10.3758/s13421-024-01603-x

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6), 110–114. https://doi.org/10.2307/3002019

Schank, R. C. (1986). *Explanation patterns: Understanding mechanically and creatively*. Erlbaum.

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metacognitive expectancy effects in source monitoring: Beliefs, in-the-moment experiences, or both? *Journal of Memory and Language, 107*, 95–110. https://doi.org/10.1016/j.jml.2019.03.009

Schmidt, O., Erdfelder, E., & Heck, D. W. (2025). How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychological Methods, 30*(4), 720–743. https://doi.org/10.1037/met0000561

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*(6), 523–542. https://doi.org/10.3758/BF03197149

Schulman, A. I. (1974). Memory for words recently classified. *Memory & Cognition, 2*(1), 47–52. https://doi.org/10.3758/BF03197491

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning & Memory, 25*(8), 369–381. https://doi.org/10.1101/lm.046912.117

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex, 19*(5), 1198–1207. https://doi.org/10.1093/cercor/bhn165

Steyvers, M., & Hemmer, P. (2012). Reconstruction from memory in naturalistic environments. *Psychology of Learning and Motivation, 56*, 25–144. https://doi.org/10.1016/B978-0-12-394393-4.00004-2

Thibault, R. T., Zavalis, E. A., Malički, M., & Pedder, H. (2024). An evaluation of reproducibility and errors in published sample size calculations performed using G*Power (p. 2024.07.15.24310458). *medRxiv.* https://doi.org/10.1101/2024.07.15.24310458

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., . . . Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied, 9*(2), 101–118. https://doi.org/10.1037/1076-898X.9.2.101

Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review, 2*(3), 387–390. https://doi.org/10.3758/BF03210977

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

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment.

Psychological Review, 90(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Unsworth, N. (2017). Examining the dynamics of strategic search from long-term memory. *Journal of Memory and Language, 93*, 135–153. https://doi.org/10.1016/j.jml.2016.09.005

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013a). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia, 51*(12), 2352–2359. https://doi.org/10.1016/j.neuropsychologia.2013.05.027

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2013b). Consolidation differentially modulates schema effects on memory for items and associations. *PLoS One, 8*(2), Article e56155. https://doi.org/10.1371/journal.pone.0056155

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences, 35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung, 18*(1), 299–342. https://doi.org/10.1007/BF02409636

Wang, T., Yue, T., & Huang, X. T. (2016). Episodic and semantic memory contribute to familiar and novel episodic future thinking. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01746

Wang, W.-C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2018). Knowledge supports memory retrieval through familiarity, not recollection. *Neuropsychologia, 113*, 14–21. https://doi.org/10.1016/j.neuropsychologia.2018.01.019

Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 687–704. https://doi.org/10.1016/S0022-5371(79)90397-9

Yacoby, A., Reggev, N., & Maril, A. (2023). Lack of source memory as a potential marker of early assimilation of novel items into current knowledge. *Neuropsychologia, 185*, Article 108569. https://doi.org/10.1016/j.neuropsychologia.2023.108569

Yonelinas, A. P. (2002). Components of episodic memory: The contribution of recollection and familiarity. In A. Baddeley, J. Aggleton, & M. Conway (Eds.), *Episodic memory: New directions in research* (pp. 31–52). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198508809.003.0003

Yonelinas, A. P., Ramey, M. M., & Riddell, C. (2024). Recognition memory: The role of recollection and familiarity. In M. J. Kahana & A. D. Wagner (Eds.), *The Oxford handbook of human memory* (pp. 923–958). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.32

Zhang, M., & Hupbach, A. (2023). The effects of variable encoding contexts on item and source recognition. *Memory & Cognition, 51*(2), 391–403. https://doi.org/10.3758/s13421-022-01353-8