

Primacy (and recency) effects in delayed recognition of items from instances of repeated events

Eva Rubínová & Heather L. Price

To cite this article: Eva Rubínová & Heather L. Price (2024) Primacy (and recency) effects in delayed recognition of items from instances of repeated events, *Memory*, 32:5, 627-645, DOI: [10.1080/09658211.2024.2354764](https://doi.org/10.1080/09658211.2024.2354764)

To link to this article: <https://doi.org/10.1080/09658211.2024.2354764>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 May 2024.



[Submit your article to this journal](#)



Article views: 168



[View related articles](#)



[View Crossmark data](#)

Primacy (and recency) effects in delayed recognition of items from instances of repeated events

Eva Rubínová^{a,b} and Heather L. Price^b

^aSchool of Psychology, University of Aberdeen, Aberdeen, Scotland; ^bDepartment of Psychology, Thompson Rivers University, Kamloops, Canada

ABSTRACT

In repeated-event paradigms where participants are asked to recall details of a sequence of similar instances they viewed/experienced previously, more accurate details are typically recalled from the first and final instances (i.e., long-term primacy and recency effects). Participants likely encode distinct attributes of details of the boundary instances that subsequently facilitate source monitoring. To date, most repeated event research has measured memory performance via free-/cued-recall paradigms; we examined delayed memory for repeated events using the recognition paradigm. In two preregistered experiments, participants viewed four videos, and after a delay completed a recognition task. In Experiment 1 ($N = 168$, between-subjects), participants decided whether an item was old (i.e., presented in any video) or new, or whether an item was presented in video 1/2/3/4 or was new. In Experiment 2 ($N = 160$, within-subjects), the old/new decision was followed by an instance attribution decision. Old items were recognised faster in the old/new task compared to the instance-attribution task. In the instance-attribution task, items from the boundary instances were accurately attributed faster compared to items from the middle instances. We found further evidence for primacy (and recency) effects in measures of confidence, memory judgments, recognition accuracy and discriminability, and confidence-accuracy calibration.

ARTICLE HISTORY

Received 5 October 2023
Accepted 7 May 2024

KEYWORDS

Repeated events; source monitoring; recognition accuracy; primacy effect; recency effect



Primacy (and recency) effects in delayed recognition of items from instances of repeated events

When remembering instances of a sequence of similar experiences (i.e., a repeated event), such as magic shows, stories describing days at a farm, or scenarios of domestic violence, people's memory for the first and final instances is typically superior to memory for the middle instances in measures of quantitative and qualitative accuracy, discriminability, rates of misattribution, and, in case of the first instance, also rates of retention (e.g., Connolly et al., 2016; Dilevski et al., 2021a; Dilevski et al., 2022; MacLean et al., 2018; Powell & Thomson, 1997; Powell et al., 2003; Roberts et al., 2015; Rubínová & Kontogianni, 2023). First, and to a lesser extent final, instances are frequently prioritised when participants choose which instance of a repeated event to report (e.g., Danby et al., 2017; cf. Dilevski et al., 2022¹). Importantly, these long-term primacy and recency effects seem to be stable across stimuli (from categorised wordlists through stories to interactive events) and delays (from 10 min to almost

two months) with primacy effects typically stronger than recency effects (Rubínová et al., 2022).

Serial position effects are also found in other areas of long-term memory. First and final instances may become landmarks in autobiographical memory serving as transitional boundaries and organising other events within these boundaries (e.g., Brown, 2016; Loftus & Marburger, 1983; Robinson, 1992; Shum, 1998; Thomsen & Berntsen, 2005). When remembering novels, films, and positions of text within a page, people recall details from the beginnings or endings more accurately than from within the boundaries (e.g., Doolen & Radvansky, 2021; Radvansky & Zacks, 2017; Rothkopf, 1971). And when judging historical events or periods, first instances and beginnings are perceived as more important, interesting, and consequential (Teigen et al., 2017).

Collectively, these findings suggest that individuals form stronger memories of details of first (and, to a lesser extent, final) experiences, and patterns of misattribution across instances of repeated events indicate stronger source memory. In other words, it appears that details

CONTACT Eva Rubínová  eva.rubinova@abdn.ac.uk  School of Psychology, University of Aberdeen, William Guild Building, Old Aberdeen AB24 3FX, Scotland

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

of the boundary instances are encoded with more unique links to instances, whereas details of the middle instances lack these links and therefore are more easily misattributed. In the present research, we focused primarily on the examination of the primacy (and recency) effect using a measure of reaction time in delayed recognition of items from instances of repeated events. We expected that stronger source memory would be evident in faster recognition judgments of presented (old) items from the first (and final) instances compared to items from the middle instances. To our knowledge, the present study is the first examination of repeated event memory using this methodology. We further examined primacy (and recency) effects more generally using indicators of accuracy and metacognition, including confidence, discriminability, and memory judgments (e.g., Tulving, 1985, 1989). This comprehensive examination enabled us to get a more holistic picture of primacy (and recency) effects in memory for repeated events. Findings of the present study may have practical implications in applied settings where individuals are interviewed about repeated events, and where interviewers decide which instance to target for reporting. Our findings may complement evidence indicating qualitative differences in memory across instances of low-frequency repeated events, particularly in terms of primacy (and recency) effects.

Mechanisms commonly used to explain serial position effects in repeated events

Ebbinghaus (1884/1964) reported that when learning sequences of syllable sets, the first set was always learned faster than subsequent sets, and the additional time oscillated around a mean value that reflected a level of fatigue and fluctuations of attention (p. 44). Ebbinghaus provides an explanation from the perspective of cognitive resources: People can pay full attention at the beginning of sequential or extended tasks when cognitive load is low, and therefore may encode first or novel experiences more easily and to more depth. James (1901) argues in a similar way about the role of attention in remembering novel and distinctive events: “The *attention* which we lend to an experience is proportional to its vivid or interesting character; and it is a notorious fact that what interests us most vividly at the time is, other things equal, what we remember best” (p. 670; emphasis in original). These mechanisms are consistent with the idea that novelty and distinctiveness drive attention and may lead to stronger encoding of first experiences (Cimbalo et al., 1978; Fabiani & Donchin, 1995; Farrar & Boyer-Pennington, 1999; Farrar & Goodman, 1990, 1992; Robinson, 1992; Teigen et al., 2017; Thomsen & Berntsen, 2005).

An attention-based mechanism is also assumed to contribute to the recency effect. Specifically, according to the schema-confirmation-deployment framework (Farrar & Boyer-Pennington, 1999; Farrar & Goodman, 1990, 1992), individuals can only pay increased attention to details of

experiences once they confirmed that the experience fits an established schema. Consequently, when experiencing a final instance of an event where an established schema can be confirmed fast, available attentional resources may enable stronger encoding of details and attributes that may uniquely link these details to the final instance (e.g., Roberts et al., 2015).

In terms of source memory, it has been demonstrated already in early research into the serial position effects that first and last items in series (be it words, wordlists, or positions of content on a page) are encoded with more contextual details (that may include positional information) than items in the middle of the series (e.g., Anderson & Bower, 1972; Bjork & Healy, 1974; Brown & Lewandowsky, 2005; Brown et al., 2007; Brown et al., 2009; Estes, 1985; Healy, 1974; Lee & Estes, 1977, 1981; Murdock, 1962; Rothkopf, 1971; Wickens, 1970). First and final experiences likely serve the role of reference points and other items in the series are encoded with information positioning them within these reference points (see gradient effects in short-term memory, Estes, 1985; Henson, 1998; and in repeated events, Rubínová & Kontogianni, 2023).

The primacy effect is frequently explained by the rehearsal of initial items while learning the rest of the list in short-term memory tasks, but rehearsal is also frequently assumed during repeated experiences that occur with longer delays apart. Specifically, when a new script is established, for example when an individual attends the first session of a language course, all event components may be initially processed as forming the event script. With repeated experience, however, some event components will be confirmed to form the script while other components will be re-evaluated as variable (e.g., Ahn et al., 1992). Rehearsal of the initial experience may occur during this process as components are compared across experiences, leading to stronger consolidation of details of the first experience (e.g., Schank, 1999; Slackman & Nelson, 1984; Underwood & Freund, 1969).

Source monitoring and measures of serial position effects in repeated events

According to the source monitoring framework, when an individual is faced with a memory retrieval task, monitoring the source of the retrieved content is an integral part of the process (Johnson et al., 1993). For example, when informing others about news, one may remember that they read part of the news online and then heard further information in a podcast. There may be a variety of attributes encoded along with the news that may help decide which information was read and which was heard. Both sources mentioned in the example are external; therefore, attributes facilitating source monitoring would likely be unique and lead to more accurate source decisions. On the other hand, internal source monitoring, such as remembering whether one wrote a practice test in class last Tuesday or the week before, would be much

more difficult. In the case of repeated events, there are substantial overlaps in many aspects of the experiences, and internal source monitoring is consequently less effective and more error prone (e.g., Lindsay, 2008; 2014).

In repeated event paradigms, where participants recall instances of a series of similar events that each contain unique details, the pattern of serial position effects in measures of accurate free recall is typically complemented by a pattern of misattributions (i.e., internal intrusions; e.g., MacLean et al., 2018; Rubínová et al., 2020). That is, participants recall a similar number of details from all instances, but they more frequently correctly attribute variable details of the boundary instances and more frequently misattribute variable details across the middle instances of repeated events. These patterns suggest lower effectiveness of source monitoring for details of middle instances, and respectively higher effectiveness of source monitoring for details of boundary instances.

Evidence indicating primacy and recency effects in repeated event paradigms comes more frequently from free- or cued-recall paradigms (e.g., Connolly et al., 2016; MacLean et al., 2018; Powell & Thomson, 1997; Powell et al., 2003; Roberts et al., 2015; Rubínová et al., 2020; 2021; 2022; Sharman et al., 2022), with few studies using a recognition paradigm (e.g., Dilevski et al., 2020; 2021a; 2022). Retrieval-level mechanisms are more strongly implicated in free-recall and cued-recall paradigms because participants need to engage in a systematic memory search to retrieve the items. These retrieval mechanisms are partially absent in recognition paradigms where participants make memory decisions for items they are presented with, although the need for involvement of retrieval processes may be greater if participants are asked questions requiring systematic source monitoring (e.g., Lindsay, 2008).

We know of two studies using the recognition task within the repeated event paradigm. Dilevski et al. (2021a) presented participants with correct and lure items and asked whether each item was seen in the designated scenario (true/false), and found recency effects for discriminability, although only in shorter delays (no delay and one week delay conditions). Dilevski et al. (2022) presented participants with items from all scenarios and asked them to decide on the source scenario. In that study, the authors did not find significant differences across instances for correct attributions and source errors (potentially due to low statistical power), but the patterns of data were consistent with the primacy and recency effects.

Present study

We presented online participants with four stories (instances of a repeated event). For stimuli sampling purposes (Wells & Windschitl, 1999), we used four sets of stimuli across participants (see Rubínová et al., 2023). Following a delay, participants were shown items from the stories intermixed with semantically related lures and

were asked to make recognition judgments. There were two types of tasks: one that required participants to retrieve source information (i.e., decide whether an item was new or presented in Instance 1, 2, 3, or 4; the instances task), and one that was free of this requirement (i.e., decide whether an item was new or old; the event task). The rationale for this manipulation was to create tasks that would vary the involvement of source monitoring and to see how this involvement of source monitoring impacted any of the measures. Because source monitoring is costly and old-new judgments (in the event task) can be made based on familiarity (e.g., McElree et al., 1999), we expected that participants' accurate recognition judgments would be faster in the event task compared to the instances task (Hypothesis 1). In addition, if items from the boundary instances of repeated events are encoded with more unique source attributes, they should be recognised faster compared to items from the middle instances. In line with findings indicating that the primacy effect is typically stronger than the recency effect (Rubínová et al., 2022), we formulated Hypothesis 2 only for the primacy effect (i.e., faster hits for items from Instance 1 than 2). Our primary hypotheses and power calculations were based on the reaction time measure, but we also measured memory judgments for items recognised as old (i.e., remember/know/not sure), decision confidence, accuracy, discriminability (i.e., the ability to differentiate old and new items), and strength of the confidence-accuracy calibration, which could provide further support for the primacy (and recency) effects.

Experiment 1

Method

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow journal article reporting standards for quantitative research in psychology (Appelbaum et al., 2018; Cooper, 2018). This study's design and its analysis were preregistered. There was one deviation from our preregistered analysis plan that is described in detail in the Statistical analysis section. Results of the preregistered analyses are reported in Online Supplemental Materials. The preregistration, data, and scripts can be accessed at <https://osf.io/hr2n3> (Rubínová & Price, 2022a). The study was approved by Thompson Rivers University Human Research Ethics Board (REB #102905).

Design

This experiment was a 2 (condition: event/instances, between-subjects) \times 4 (instance: 1/2/3/4, within-subjects) mixed design. We measured reaction time and confidence for recognition decisions and memory judgments for items recognised as old (remember/know/not sure). We

constructed receiver-operating characteristic (ROC) curves to evaluate discriminability and confidence-accuracy (CA) curves to evaluate calibration across instances.

Participants

Power analysis. We calculated the required sample using a simulation-based power analysis (package Superpower; Lakens & Caldwell, 2021) for a 2 (between-subjects) \times 4 (within-subjects) mixed design. We estimated means of reaction times corresponding to a large effect of condition, Cohen's $f = .47$ and a medium effect of instance, Cohen's $f = .13$, used the default value of within-subject correlation, $r = .5$, and set $\alpha = .03$ and desired power around 90%.¹ The required sample size for the between-subjects effect of condition was 29 and for the within-subjects effect of instance was 162. To counterbalance our stimuli, we needed to achieve the final sample in multiples of 16; therefore, we planned to recruit 160 participants for this study (power = 89.77%).³

Inclusion criteria. Participants were recruited via Prolific, an online platform where vetted participants can take part in paid online studies, using the following screeners validated within the experiment: age between 18 and 55 years and self-reported English language fluency (we validated these screeners in self-report question at the beginning of the study).

Exclusion criteria. To ensure that (1) lack of attention during stimuli presentation and (2) lack of motivation during the memory task would not impact the results, we excluded data from participants who indicated that they paid low attention and/or had low motivation (ratings < 10% on a 0-100% scale). We also excluded data from participants who reported that they had issues viewing any of the stimuli videos. For reaction time values, in line with our pre-registered plan, we excluded outliers that were defined as values exceeding 2 SDs from the group means for each instance in each condition (Berger & Kiefer, 2021; Morís Fernández & Vadillo, 2020). Two percent of trials were excluded following this procedure, and there was marked improvement in skewness (20.79–2.29) and kurtosis (747.74–6.73). To further improve the distribution of reaction time data, we used power transformation from the `car` package (Fox & Weisberg, 2018), and arrived at further improved skewness (0.03) and kurtosis (1.62).

Sample. Due to uncertainty in numbers of exclusions, we oversampled and recruited 197 participants in total from Prolific. We excluded 21 participants who reported having issued while viewing the videos, 6 participants who reported low motivation during the recognition task, and 2 participants who reported low attention during stimuli presentation. The final sample consisted of 168 participants. Demographics were obtained from Prolific, and our final sample included 108 females and 57 males (data from 3 participants are not available) aged between 19 and 55 years ($M = 35.21$, $SD = 10.10$, data from 3 participants are not available but these

participants confirmed they were between 18 and 55 years of age). Participants reported White ($N = 100$), Asian ($N = 24$), Black ($N = 16$), Mixed ($N = 14$), and other ($N = 7$) race (data from 7 participants are not available).

Materials

We used four sets of stimuli in the presentation phase. Each set contained four videos (1.25-1.55 min) including narrative stories adapted from Rubínová et al., 2022. Set 1 comprised four variations of a birthday party preparation; Set 2 followed a creature building a machine; Set 3 depicted a group of people planting spying devices at various locations; and Set 4 described four days at a farm. For each set, participants were presented with the four videos in one of four orders (ABCD, BCDA, CDAB, DABC); therefore, there were 16 versions of the experiment in each condition. Each video contained 15 unique items (story details; e.g., "Green Moon"; "Hot Chocolate"), and when mentioned in the narrative, these items appeared at the bottom of the screen in large font for a duration of 2 s. In the recognition phase, participants were randomly presented with 60 old items (15 items from each video) and 60 new items (not presented semantically related items that were plausible alternatives to presented items).⁴

Procedure

Following consent, sound check, and inclusion criteria screeners, participants were presented with the following instructions: "You will be presented with a series of videos about **Emma**. There will be 2 min of mathematical tasks after each video. The mathematical tasks are timed and the screen will automatically advance once the time limit is reached. **Please pay attention to the videos and the highlighted words.** Each video will start playing automatically, will be played only once, and you will not be able to pause any of the videos. Click **Continue** once ready to start." Note that videos were either about preparation of a birthday party, T-44, outdoor scenes, or Emma; assignment to stimuli was counterbalanced. Participants then viewed four videos (Figure 1). After each video, participants were asked how much attention they paid to it (0-100% sliding scale labelled "Low attention" and "High attention" at the extremes). Two-minute arithmetic filler tasks were administered between the videos. After the final video, participants completed a 10-minute filler task (they viewed and rated similarity and understanding of four unrelated videos).

Participants were then randomly assigned to the event condition ($N = 85$) or the instances recognition condition ($N = 83$) and we aimed at approximately counterbalanced allocation across the 32 combinations of stimuli and condition. The final allocation in the event condition: $N_{Set\ 1} = 21$, $N_{Set\ 2} = 21$, $N_{Set\ 3} = 22$, $N_{Set\ 4} = 21$; and in the instances condition: $N_{Set\ 1} = 26$, $N_{Set\ 2} = 20$, $N_{Set\ 3} = 17$, $N_{Set\ 4} = 20$. Note that we intended high stimulus variability to increase

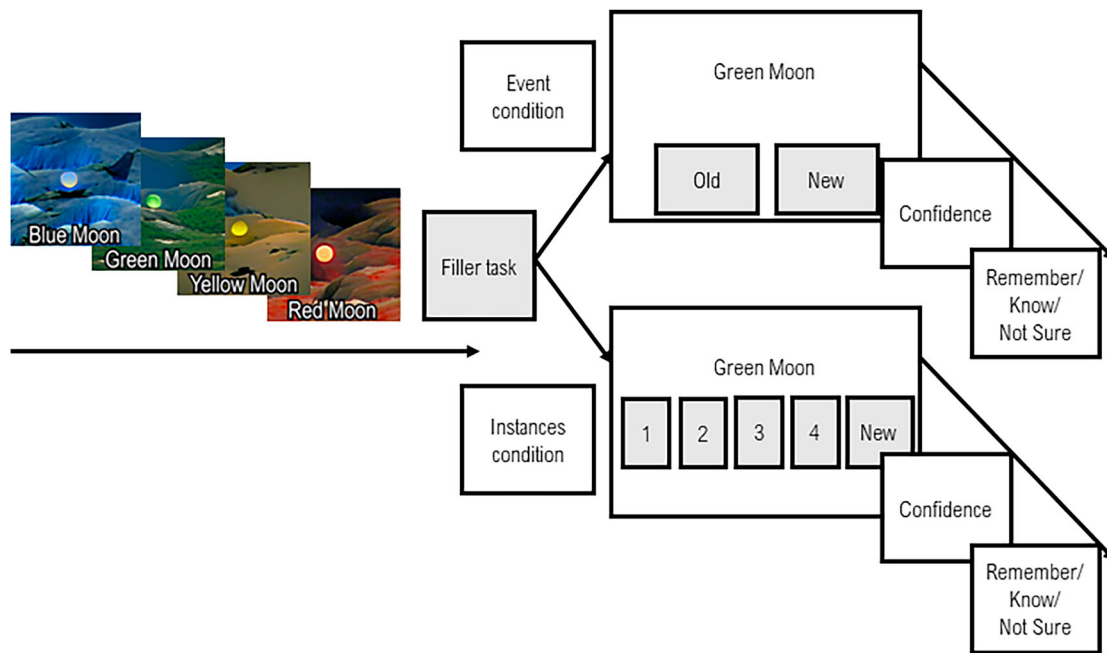


Figure 1. Flowchart of the procedure in Experiment 1.

generalisability of our findings, and we addressed variability at the level of stimuli by using random intercepts for individual words in statistical analyses (see below).

In the event condition, participants were presented with items and asked whether the item was Old (presented within one of the videos) or New (not presented in any of the videos). In the instances condition, participants were told to decide whether the item was presented in Video 1, 2, 3, 4, or was New (not presented in any of the videos). After each decision, participants were asked to report their confidence (0-100% sliding scale labelled “Low confidence” and “High confidence” at the extremes). For all Old items (event condition) or all items not marked as New (instances condition), participants were asked to provide memory judgment responses: “Do you **Remember** the word being in one of the videos? OR Do you just **Know** it was in one of the videos (without remembering it)?” with the following response options: Remember, Know, and Not Sure.

Following the recognition task, we asked participants how motivated they were to complete the task (0-100% sliding scale labelled “Low motivation” and “High motivation” at the extremes), and whether they experienced any issues viewing the videos (Yes/No).

Measures

Reaction time. We measured time from the onset of item presentation until participants clicked a response button.

Confidence. Participants indicated their level of confidence associated with a decision they made on a 0-100% slider scale.

Memory judgments. For each item recognised as occurring in one of the videos, participants stated

whether they remembered the item, simply knew the item was old, or were not sure.

Recognition decisions. In the event condition, participants either accurately recognised an old item (hit), correctly stated that a new item was new (correct rejection) incorrectly stated that an old item was new (miss), or incorrectly stated that a new item was old (false alarm). In the instances condition, participants either recognised an old item and accurately attributed it to the instance in which it occurred (hit), recognised an old item and inaccurately attributed it to an instance in which it did not occur (misattribution), incorrectly stated that an old item was new (miss), or correctly stated that a new item was new (correct rejection).

ROC curves. To evaluate decision accuracy, we constructed confidence-based and reaction time-based ROC curves for each instance and condition (Brady et al., 2023). For confidence, we first split decisions into 11 confidence bins (bin 1 = 100% confidence; bin 2 = confidence 90-99%; bin 3 = confidence 80-89%; up to bin 11 = 0% confidence). To create 11 points for the ROC curve, we computed cumulative hit, false alarm, and misattribution rates. The hit rate for the first point was computed by dividing the sum of hits for confidence bin 1 by the number of old trials. For the second point, we added hits for confidence bins 1 and 2 and divided them by the number of old trials. Hit rates for the rest of the points were computed by adding hits associated with the next confidence bin and dividing them by the number of old trials. False alarm rates were computed analogically for false alarms among new trials. In the instances condition, we computed misattribution rates based on misattributions within old trials. For conciseness, and due to

consistency across results, we report reaction time-based ROC curves in the Online Supplemental Materials.

CA calibration curves. To calculate accuracy in the event condition, we divided the number of hits and correct rejections by the sum of hits, correct rejections, misses, and false alarms for each confidence bin. To calculate accuracy in the instances condition, we divided the number of hits and correct rejections by the sum of hits, correct rejections, misses, false alarms, and misattributions for each confidence bin.

Statistical analyses

To assess our primary hypotheses, we computed a linear mixed model (LMM) with reaction time as the dependent variable, condition and instance and their interaction as fixed factors, and with random slopes for instance and random intercepts for subjects and stimuli items. Condition and instance were coded with successive difference contrasts (Schad et al., 2020). For condition, the contrast compared event and instances groups. For instance, the contrasts compared reaction times between items from Instance 1 and 2 (the primacy effect), items from Instance 2 and 3, the items from Instance 3 and 4 (the recency effect). This analytical approach enabled us to estimate subject- and stimuli-level effects. We believe that analysing data at the level of item responses is appropriate for our data, although it is a deviation from our preregistered analysis plan, where we set out to analyse averaged reaction times per participant and instance. Results of these analyses are reported in the Online Supplemental Materials and are consistent with results reported in the main text.

Confidence data were analysed the same way as reaction time only with confidence ratings as the dependent variable. For memory judgments, we built three generalised LMMs to evaluate the odds of judging items as remembered, known, or not sure between conditions and across instances.

Accuracy was evaluated in four ways. First, we assessed differences in recognition decisions between conditions and across instances in three generalised LMMs: hits/misses, correct rejections/false alarms, and, in the instances condition, hits/misattributions. Next, we constructed confidence-based and reaction-time based ROC curves for hit and false alarm rates (both conditions) and hit and misattribution rates (instances condition). For each set of ROC curves, the topmost curve would indicate better discriminability as it shows higher hit rate at the same level of false alarm/misattribution rate. Finally, to assess the strength of the relationship between confidence and accuracy in the two conditions and across instances, we constructed confidence-accuracy calibration curves.

To correct for Type I error, we computed a boundary value for each family of analyses using the false discovery rate (FDR) correction approach (Benjamini & Hochberg, 1995). A family of analyses was defined as all tests related to a measure, e.g., all 13 tests evaluated for reaction time analyses of hits. The FDR was calculated for each rank

of the 13 ranked p -values by dividing the rank by the number of tests and multiplying this value by alpha (.05). The boundary value is then the first FDR value that is greater than the p -value at the corresponding rank (i.e., all p -values < this FDR are considered significant; all p -values in higher ranks are considered nonsignificant; all FDR calculations are provided on the OSF). Data were analysed and visualised in R (R Core Team, 2020) using packages ggplot2 (Wickham, 2016), lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), MASS (Venables & Ripley, 2002), psych (Revelle, 2023), reshape2 (Wickham, 2007), and stringr (Wickham, 2022).

Results and discussion

Reaction time

Old items were accurately recognised faster in the event condition compared to the instances condition, $b = -0.66$, $[-0.89, -0.42]$, $t = 5.50$, $p < .001$ (preregistered Hypothesis 1). There were also differences in hit latencies across instances (Table 1; Figure 2). Items from Instance 1 were recognized and accurately attributed faster than items from Instance 2 (preregistered Hypothesis 2), $b = -0.18$, $[-0.24, -0.07]$, $t = 3.59$, $p < .001$; hits for items from Instance 2 were faster than hits for items from Instance 3, $b = -0.12$, $[-0.21, -0.02]$, $t = 2.42$, $p = .017$; and hits for items from Instance 4 were also faster than hits for items from Instance 3, $b = -0.12$, $[-0.21, -0.03]$, $t = 2.54$, $p = .012$. All instance-level differences showed an interaction with condition ($ps \leq .030$). In the following section, we provide results of analyses split by condition, which indicated that these effects were mainly driven by the instances condition.

In the instances condition, participants recognized and accurately attributed items from Instance 1 faster than items from Instance 2, $b = -0.22$, $[-0.38, -0.05]$, $t = 2.54$, $p = .013$, and also items from Instance 2 than 3, $b = -0.21$, $[-0.38, -0.04]$, $t = 2.39$, $p = .019$. The recency effect was not significant following the FDR correction, $b = -0.20$, $[-0.40, -0.01]$, $t = 2.01$, $p = .048$. There were no significant instance effects in the event condition ($ps \geq .412$).

There were no significant differences in reaction times for misses and correct rejections ($ps \geq .078$), and there were also no significant differences in reaction times for misattributions in the instances condition ($ps \geq .090$). Decisions leading to false alarms were faster in the event condition compared to the instances condition, $b = -0.45$, $[-0.72, -0.18]$, $t = 3.22$, $p = .002$.

Confidence

Participants were overall more confident when they accurately recognized items in the event condition than in the instances condition, $b = 14.25$, $[9.72, 18.78]$, $t = 6.16$, $p < .001$ (Table 1). There were also differences consistent

Table 1. Reaction time (non-transformed) and confidence ratings across recognition decisions, conditions, and instances in Experiment 1.

Variable/ Level	Hit		Miss		Correct rejection		False alarm		Misattribution	
	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N
Reaction time (milliseconds)										
Event										
1	2403 (1806)	910	2553 (2092)	344	2579 (2019)	850	2623 (1921)	391	–	–
2	2428 (1572)	852	2382 (1800)	391	2522 (1720)	884	2538 (1629)	356	–	–
3	2509 (1856)	835	2633 (2388)	428	2824 (2299)	884	2876 (2383)	372	–	–
4	2395 (1513)	823	2357 (1685)	423	2632 (1796)	893	2820 (1969)	341	–	–
Total	2433 (1696)	3420	2480 (2013)	1586	2640 (1973)	3511	2713 (1999)	1460	–	–
Instances										
1	3426 (1500)	402	3018 (2542)	271	2975 (2436)	719	3367 (2603)	483	3423 (2605)	527
2	3774 (2499)	364	2899 (2271)	308	2967 (2360)	739	3410 (2592)	465	3442 (2468)	524
3	4779 (3594)	364	2975 (2710)	351	2976 (2579)	749	3366 (2694)	466	3705 (2932)	513
4	3866 (2605)	374	2758 (1859)	354	2788 (2081)	737	3315 (2323)	461	3242 (2013)	465
Total	3947 (2869)	1504	2906 (2358)	1284	2927 (2372)	2944	3365 (2556)	1875	3457 (2541)	2029
Confidence										
Event										
1	79.80 (22.95)	921	54.98 (26.52)	349	60.64 (27.46)	867	61.17 (26.73)	397	–	–
2	76.42 (24.64)	869	55.26 (26.85)	400	61.65 (26.37)	900	59.37 (27.07)	364	–	–
3	77.55 (23.62)	839	53.07 (26.47)	431	61.26 (26.80)	888	58.53 (26.33)	376	–	–
4	77.99 (23.55)	835	53.37 (26.69)	435	62.06 (27.23)	913	60.80 (27.26)	351	–	–
Total	77.97 (23.71)	3464	54.11 (26.63)	1615	61.42 (26.96)	3568	59.98 (26.83)	1488	–	–
Instances										
1	76.83 (26.11)	413	47.56 (32.39)	279	53.23 (33.21)	735	46.33 (27.77)	502	51.30 (28.14)	548
2	59.71 (29.41)	382	48.77 (32.93)	321	53.50 (33.73)	752	46.63 (26.82)	484	56.56 (29.96)	537
3	59.45 (29.25)	364	45.36 (32.68)	353	51.26 (33.38)	759	45.43 (27.29)	477	55.86 (29.91)	523
4	67.81 (27.32)	398	46.84 (33.35)	362	53.65 (34.11)	754	46.42 (28.06)	483	52.72 (27.57)	480
Total	66.26 (28.88)	1557	48.07 (32.85)	1315	52.91 (33.61)	3000	46.21 (27.48)	1946	54.12 (29.00)	2088

Note. *N*s for reaction time and confidence differ due to exclusions of outlier reaction time values (~2% of data).

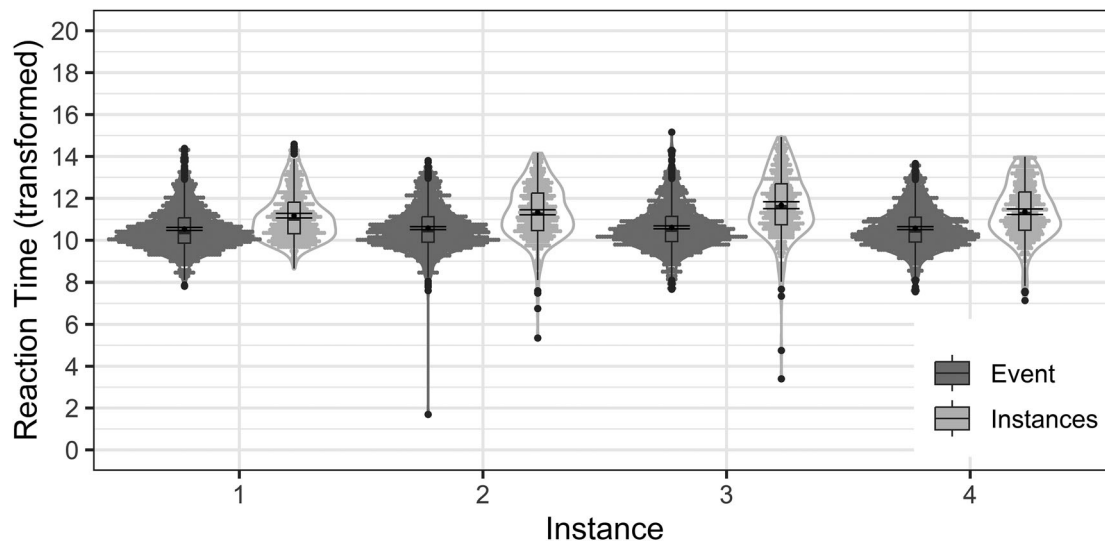


Figure 2. Hit response latencies (transformed) across instances and conditions in Experiment 1.

Note. Points represent individual response latencies per participant and instance following preregistered exclusion of values $\pm 2SD$ and power transformation. Box plots present medians and upper and lower quartiles. Black points with error bars present means and 95% CIs, and black points at the top of the distributions represent outlier values.

with the primacy and recency effects: participants were more confident when accurately recognising items from Instance 1 than items from Instance 2, $b = 10.06$, [7.96, 12.15], $t = 9.42$, $p < .001$, and items from Instance 4 than items from Instance 3, $b = 3.37$, [1.26, 5.49], $t = 3.13$, $p = .002$. Both instance effects were qualified by significant interactions: the primacy effect was more pronounced in the instances condition, $b = 17.93$, [13.58, 22.28], $t = 8.08$, $p < .001$, than in the event condition, $b = 3.40$, [1.37, 5.43], $t = 3.29$, $p = .001$; and the recency effect was only present in the instances condition, $b = 7.27$, [4.03, 10.52], $t = 4.42$, $p < .001$, not in the event condition, $p = .545$.

There were no significant effects for confidence judgments of missed items, correct rejections, and misattributions after the FDR corrections ($ps \geq .016$). For false alarms, participants were more confident in the event than the instances condition, $b = 15.24$, [9.83, 20.66], $t = 5.52$, $p < .001$.

Memory judgments

Participants in the event condition judged items more often as remembered compared to the instances condition, $OR = 2.05$, [1.41, 2.99], $z = 3.73$, $p < .001$ (Table 2). Participants also judged items from Instance 1 more frequently as remembered compared to items from Instance 2, $OR = 1.25$, [1.07, 1.46], $z = 2.81$, $p = .005$. Inverse to this pattern were not sure judgments: participants in the

instances condition more frequently reported that they were not sure compared to the event condition, $OR = 3.11$, [2.00, 4.84], $z = 5.04$, $p < .001$. Participants also more frequently reported that they were not sure about their memory for items from Instance 2 than 1, $OR = 1.32$, [1.09, 1.60], $z = 2.79$, $p = .005$. There were no further significant effects or interactions ($ps \geq .173$), and no differences in judging items as known ($ps \geq .322$).

Accuracy

Recognition decisions. Table 3 presents proportions of recognition decisions across instances and conditions. To create a comparable measure of hits as an indicator of recognition of old items in both conditions, we treated misattributions in the instances condition as hits (i.e., these were items recognized as old but attributed to incorrect instances). Old items were accurately identified more often in the instances condition than in the event condition, $OR = 1.43$, [1.08, 1.90], $z = 2.49$, $p = .013$, and there were more hits for items from Instance 1 than for items from Instance 2, $OR = 1.32$, [1.09, 1.59], $z = 2.89$, $p = .004$. In new trials, there were more false alarms in the instances than event condition, $OR = 1.34$, [1.10, 1.64], $z = 2.85$, $p < .001$. There were no differences in misattribution rates in the instances condition and no further significant effects or interactions ($ps \geq .054$).

Table 2. Proportions of memory judgments across conditions, instances, and experiments.

Judgment	Experiment 1								Experiment 2 (Instances Only)			
	Event				Instances							
	1	2	3	4	1	2	3	4	1	2	3	4
Remember	.65	.63	.64	.64	.53	.49	.47	.51	.62	.58	.57	.57
Know	.22	.21	.22	.23	.21	.24	.22	.23	.22	.22	.23	.23
Not Sure	.12	.15	.14	.13	.26	.28	.30	.27	.17	.20	.20	.20

Table 3. Proportions of recognition decisions across conditions and instances in Experiment 1.

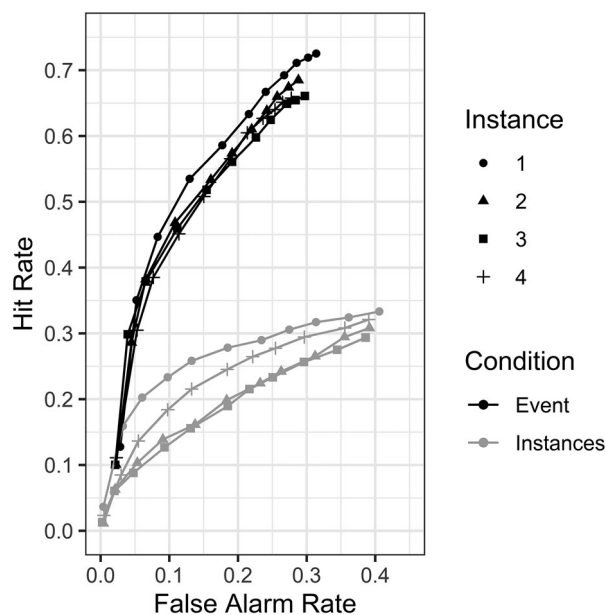
Decision	Condition							
	Event				Instances			
	1	2	3	4	1	2	3	4
Old trials								
Hit	.73	.68	.66	.66	.33	.31	.29	.32
Miss	.27	.32	.34	.34	.22	.26	.28	.29
Misattribution	–	–	–	–	.44	.43	.42	.39
New trials								
Correct rejection	.69	.71	.70	.72	.59	.61	.61	.61
False alarm	.31	.29	.30	.28	.41	.39	.39	.39

Confidence-based ROC curves. Figure 3 displays ROC curves constructed from hit and false alarm rates at 11 levels of confidence for each instance and condition. In the event condition, discriminability reflects higher accuracy of identifying old items; in the instances condition, discriminability reflects higher accuracy of source attribution decisions. Therefore, the ROC curves are incomparable between conditions, but they allow us to compare discriminability across instances. Curves that are consistently above other curves indicate better discriminability as they show higher hit rates at the same level of false alarm rates.

The topmost ROC curves in both conditions correspond to performance for items from Instance 1, suggesting better discriminability between old and new items, and in the instances condition, also more accurate source attribution, compared to items from other instances. This primacy effect was stronger in the instances condition, where there was additionally an indication of the recency effect: the ROC curve associated with items from Instance 4 was consistently above the ROC curves for items from Instances 2 and 3.

The left panel of Figure 4 displays ROC curves constructed from hit and misattribution rates in the instances condition that indicate accuracy of source attribution decisions. The topmost ROC curve indicates higher discriminability for items from Instance 1 (primacy effect). The second highest ROC curve associated with items from Instance 4 is consistently above the ROC curves associated with items from Instances 3 and 4 (recency effect).

CA calibration curves. Figure 5 displays CA calibration curves for each instance in both conditions. In both conditions, the curves indicate good confidence-accuracy calibration only at very low levels of confidence: <40% in the event and <20% in the instances condition, and overconfidence beyond these levels of confidence. In the event task, the curves associated with items from the four instances overlap and therefore there is no indication of differences in calibration. In the instances task, there is high variability, and the only clear effect seems to be better confidence-accuracy calibration for items from Instance 1 at high levels of confidence (<89%) that is consistent with the primacy effect.

**Figure 3.** Confidence-based receiver-operating characteristic curves for hit and false alarm rates across instances and between conditions in Experiment 1.

Summary

In terms of differences between conditions, we found support for our preregistered Hypothesis 1: old items were recognized faster in the event condition than in the instances condition. Compared with the instances condition, participants in the event condition also more frequently judged old items as remembered, were less frequently unsure about why they judged items as old, reported higher confidence when accurately recognising old items, and showed stronger confidence-accuracy relationship. These condition-level differences are not surprising, given that decisions in the event task were easier to make as participants only needed to detect items that were familiar (e.g., McElree et al., 1999). It was only in the instances condition that participants needed to engage source monitoring and make attribution decisions, which likely increased decision time, led to lower confidence, and greater uncertainty in memory judgments as participants would only judge items as remembered when they remembered them along with the source instance.

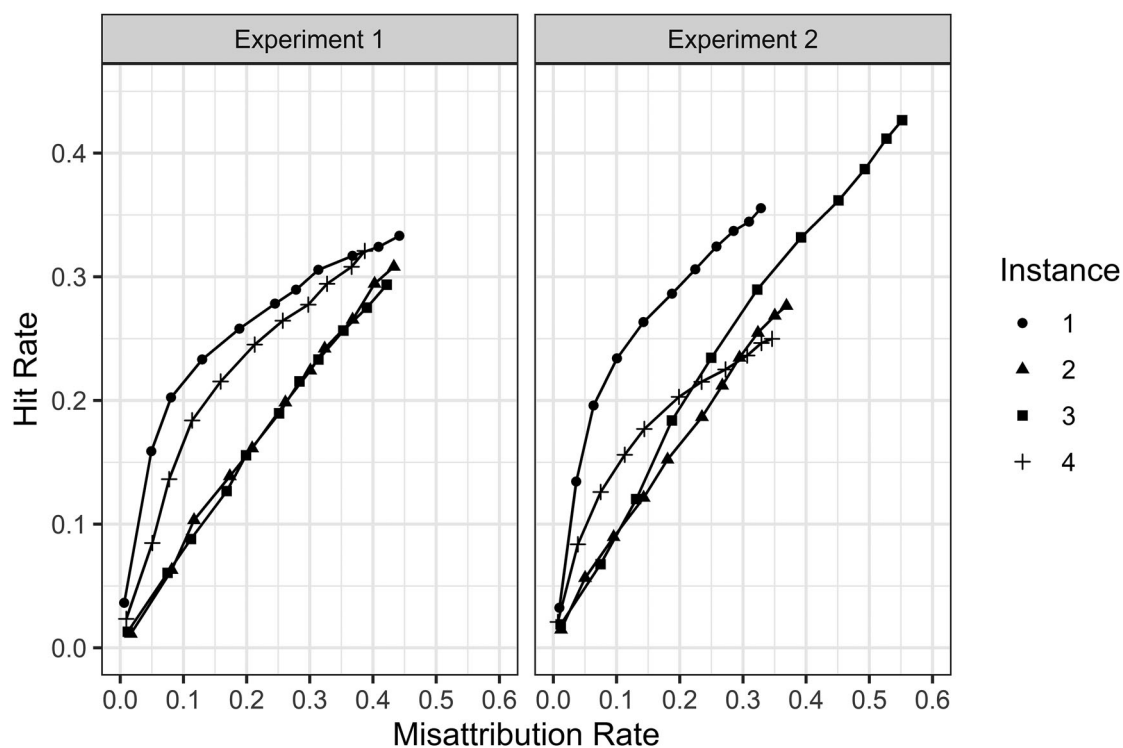


Figure 4. Confidence-based receiver-operating characteristic curves for hit and misattribution rates (instances condition/task) across instances in Experiments 1 and 2.

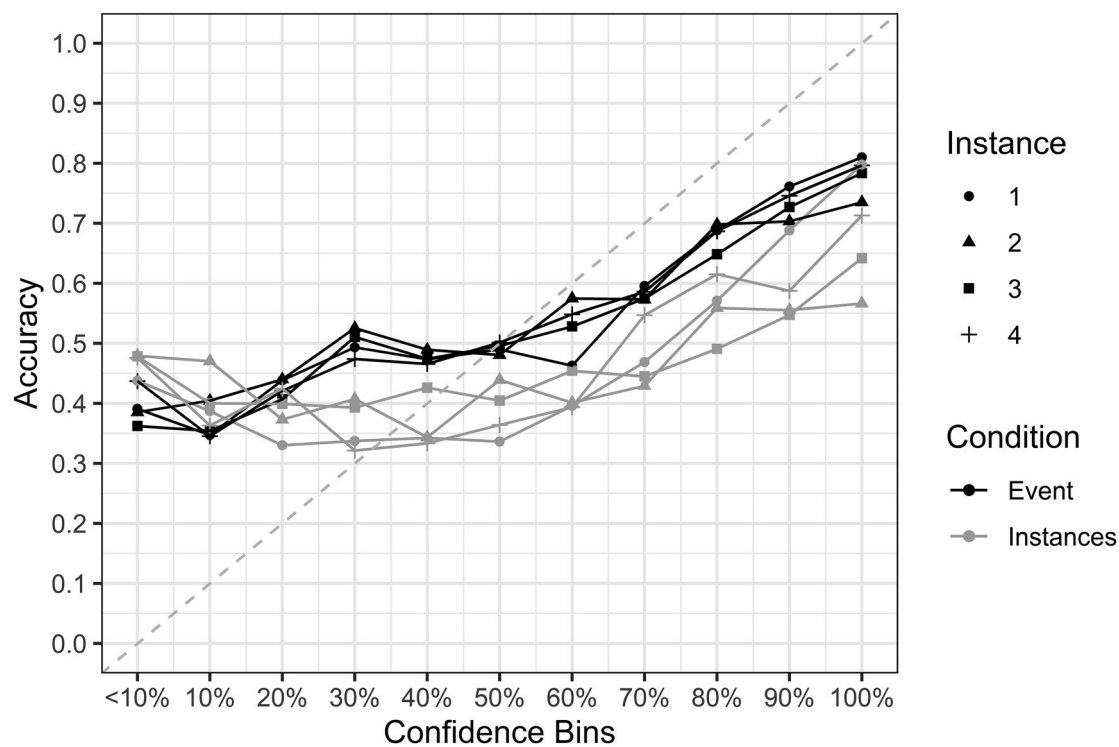


Figure 5. Confidence-accuracy calibration curves between conditions and across instances in Experiment 1.

Note. Dashed reference line represents perfect calibration.

In terms of differences across instances, we found partial support for our preregistered Hypothesis 2: in the instances condition – but not in the event condition – recognition of old items was faster for Instance 1 than Instance 2 (i.e., the primacy effect). Support for the primacy effect was found in memory judgments. In both conditions, old items from Instance 1 were more frequently judged as remembered and less frequently judged unsure compared to old items from Instance 2. Further support for the primacy effect was found in confidence ratings: hits for items from Instance 1 were associated with higher confidence compared to hits for items from Instance 2, and although more pronounced in the instances condition, this effect was present in both conditions. Finally, the primacy effect also emerged in accuracy analyses: participants recognized more old items from Instance 1 than 2, and ROC curves indicated consistently higher discriminability for items from Instance 1 (compared to other instances) in both conditions. Note that we found no differences across instances for false alarms and correct rejections – these differences would not be expected as they concerned new items and therefore cannot be impacted by any differences in encoding. Analyses of the confidence-accuracy relationship indicated better calibration for items from Instance 1 (compared to other instances) at high levels of confidence (> 85% confidence).

There was additionally an indication of the recency effect. Participants recognized old items from Instance 4 faster than old items from Instance 3, and they also indicated higher confidence (compared to items from Instance 3) and source discriminability (compared to items from Instances 2 and 3). At high levels of confidence, confidence-accuracy calibration was better for items from Instance 4 than items from Instances 2 and 3. We did not find support for the recency effect in analyses of memory judgments or recognition decisions.

We additionally found a significant difference in response latencies for accurately recognized items between the two middle instances (responses were faster for items from Instance 2 than Instance 3). We had no expectations of this effect, although sometimes significant differences in accuracy are found between the middle instances (e.g., Rubínová et al., 2022). Instead of speculating about why these differences emerged, we wanted first to see if they replicate.

Experiment 2

In Experiment 2, we intended to replicate findings from Experiment 1. To better isolate the reaction time associated with decisions that lead to instance attribution after recognising an old item, we designed Experiment 2 as fully within-subjects: participants first engaged in old/new recognition (the event task) and then, for items recognized as old, engaged in source attribution decision (the instances task). Although we formulated a hypothesis related to task in the preregistration (parallel to

Experiment 1), given the tasks occurred in sequence, we focused primarily on our hypotheses regarding the primacy and recency effects.

Method

Experiment 2 used the same design and measures as Experiment 1 except for condition, which was turned into task (event/instances) and administered within-subjects. Specifically, following the presentation phase and filler task, participants were presented with an item and first made an old/new judgment. For all items judged new, participants then completed a confidence rating, and the next item was presented. For all items judged old, participants were asked to decide whether the item was presented in Video 1, 2, 3, or 4 (i.e., the instance attribution decision), rate their confidence, and complete a remember/know/not sure judgment rating before the next item was presented.

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow journal article reporting standards for quantitative research in psychology (Appelbaum et al., 2018; Cooper, 2018). This study's design and its analysis were preregistered. As in Experiment 1, we deviated from our preregistered analysis by not averaging reaction time data at the level of participants (see Statistical analysis section). Results of the preregistered analyses are reported in Online Supplemental Materials. The preregistration, data, and scripts can be accessed at <https://osf.io/hr2n3> (Rubínová & Price, 2022b). The study was approved by Thompson Rivers University Human Research Ethics Board (REB #102905).

Participants

Our target sample was 160 participants (for power analysis, see Experiment 1) and we recruited 175 participants in total. Eleven participants were excluded due to reporting issues when viewing the stimuli and 4 participants were excluded because they reported low motivation during the recognition task. The final sample with demographic information obtained from Prolific consisted of 160 participants (103 females and 55 males, data from 2 participants are not available) aged between 19 and 56 years ($M = 33.16$, $SD = 9.44$). Participants reported White ($N = 111$), Black ($N = 14$), Mixed ($N = 13$), Asian ($N = 12$), and other ($N = 8$) race; data from 2 participants are not available.

As in Experiment 1, we excluded reaction time values exceeding 2 SD s from the group means for each instance in each task (Berger & Kiefer, 2021; Morís Fernández & Vadillo, 2020). Two percent of trials were excluded and there was an improvement in skewness (33.62–3.14) and kurtosis (2156.82–13.81). The distribution further improved following power transformation: final skewness = -0.01 and kurtosis = 0.32.

Results and discussion

Reaction time

In the analysis of hit latencies, we again found evidence for the primacy and recency effects. Items from Instance 1 were accurately recognized faster than items from Instance 2 (pre-registered Hypothesis 2), $b = -0.05$, $[-0.06, -0.03]$, $t = 7.76$, $p < .001$, and items from Instance 4 were accurately recognized faster than items from Instance 3 (preregistered Hypothesis 2), $b = -0.02$, $[-0.3, -0.01]$, $t = 4.03$, $p < .001$ (Table 4). There was also a difference between items from the middle instances: items from Instance 2 were accurately recognized faster than items from Instance 3, $b = -0.01$, $[-0.03, -0.003]$, $t = 2.57$, $p = .010$.

As indicated by a significant interaction with task ($ps \leq .001$), the primacy and recency effects were stronger in the instances task, primacy: $b = -0.09$, $[-0.11, -0.07]$, $t = 7.42$, $p < .001$; recency: $b = -0.04$, $[-0.06, -0.01]$, $t = 2.94$, $p = .004$, and were not significant in the event task ($ps \geq .044$).

There were no significant differences in latencies for misses, false alarms, correct rejections in the event task, or misattributions in the instances task after the FDR correction ($ps \geq .022$).

Confidence

For hits in the instances task, participants were more confident in their decisions for items from Instance 1 than 2 (primacy effect), $b = 12.94$, $[10.04, 15.85]$, $t = 8.73$, $p < .001$, and for items from Instance 4 than 3 (recency effect), $b = 8.94$, $[6.07, 11.80]$, $t = 6.12$, $p < .001$ (Table 4). There were no significant differences in confidence ratings for misses, false alarms, and correct rejections in the event task ($ps \geq .205$). For misattributions in the instances task, confidence ratings were lower for misattributed items from Instance 1 than 2, $b = -3.27$, $[-1.03, -5.51]$, $t = 2.86$, $p = .005$.

Memory judgments

Memory judgments were evaluated for each item following instance attribution decision in the instances task. We found evidence for the primacy effect, with items from Instance 1 more frequently judged as remembered compared to items from Instance 2, $OR = 1.27$, $[1.06, 1.51]$, $z = 2.61$, $p = .009$ (Table 2). Participants also more frequently reported that they were not sure about their memory for items from Instance 2 than 1, $OR = 1.45$, $[1.08, 1.94]$, $z = 2.50$, $p = .013$. There were no differences for known judgments and no further significant effects ($ps \geq .324$).

Accuracy

Recognition decisions. Table 5 presents proportions of recognition decisions across tasks and instances. For old

items in the event task, there were more hits (than false alarms) for items from Instance 1 than 2, $OR = 1.35$, $[1.14, 1.61]$, $z = 3.44$, $p < .001$, and there were more hits for items from Instance 3 than 4, $OR = 1.20$, $[1.03, 1.40]$, $z = 2.28$, $p = .023$. For old items in the instances task, there were more hits (than misattributions) for items from Instance 1 than 2, $OR = 1.50$, $[1.21, 1.85]$, $z = 4.47$, $p < .001$. There were no further significant effects after the FDR correction ($ps \geq .035$). As is apparent from Table 5, there were almost no differences in false alarms for new items across instances in the event task ($ps \geq .422$).

Confidence-based ROC curves. The right panel of Figure 4 displays ROC curves constructed from hit and misattribution rates in the instances task that indicate accuracy of source attribution decisions. The topmost ROC curve indicates higher accuracy of source attribution decisions for items from Instance 1 (primacy effect). There is no further evidence of consistent differences in the accuracy of decisions for items from other instances (i.e., the curves are crossed).

CA calibration curves. Figure 6 displays CAC curves from the instances task and therefore represent calibration between source memory decisions and confidence. The curves indicate poor confidence-accuracy calibration up to 60% confidence (the curve is flat), and there seems to be better calibration for higher levels of confidence with overall overconfidence. Consistently with findings from Experiment 1, at higher levels of confidence ($>79\%$), there is better confidence-accuracy calibration for items from Instance 1 compared to other instances indicating the primacy effect. In addition, there is also an indication of the recency effect – better calibration for items from Instance 4 compared to Instances 2 and 3 at higher levels of confidence ($>79\%$).

Summary

We found support for primacy and recency effects in reaction time analyses (preregistered Hypothesis 2), ratings of recognition decision confidence, and confidence-accuracy calibration at high levels of confidence ($> 70\%$, with a stronger primacy effect). Further support for the primacy (but not recency) effect was found in analyses of memory judgments, recognition decision, and ROC curves. Finally, indirect support for the primacy effect was found in misattributions: participants were less confident in misattribution decisions associated with items from Instance 1.

Discussion

In the present study, we examined primacy (and recency) effects in recognition of items from instances of repeated events in measures of reaction time, confidence, indicators of accuracy and discriminability, and metacognitive indicators including judgments of memory and the confidence-accuracy relationship. We found that when

Table 4. Reaction time (non-transformed) and confidence ratings across recognition decisions, tasks, and instances in Experiment 2.

Variable/ Level	Hit		Miss		Correct rejection		False alarm		Misattribution	
	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N
Reaction time (milliseconds)										
Event										
1	2589 (1955)	1605	2528 (2210)	745	2466 (2047)	1742	2764 (2218)	600	–	–
2	2699 (2210)	1523	2550 (2392)	836	2588 (2267)	1740	2845 (2291)	609	–	–
3	2902 (2359)	1490	2347 (2102)	861	2570 (2330)	1758	2920 (2484)	589	–	–
4	2856 (2422)	1414	2462 (2452)	952	2658 (2616)	1774	3161 (2850)	581	–	–
Total	2758 (2239)	6032	2469 (2300)	3394	2571 (2326)	7014	2920 (2472)	2379	–	–
Instances										
1	1512 (1365)	820	–	–	–	–	–	–	1798 (1489)	765
2	1972 (1608)	628	–	–	–	–	–	–	1723 (1341)	852
3	2045 (1615)	620	–	–	–	–	–	–	1801 (1464)	821
4	1841 (1496)	573	–	–	–	–	–	–	1632 (1258)	795
Total	1818 (1528)	2641	–	–	–	–	–	–	1738 (1391)	3233
Confidence										
Event										
1	–	–	51.62 (28.39)	755	56.77 (29.93)	1771	47.76 (27.26)	610	–	–
2	–	–	50.16 (29.16)	848	55.54 (30.20)	1759	48.33 (26.01)	621	–	–
3	–	–	51.63 (27.66)	883	56.63 (29.86)	1777	46.39 (25.54)	602	–	–
4	–	–	51.46 (28.85)	966	56.16 (29.79)	1788	45.30 (26.05)	592	–	–
Total	–	–	51.22 (28.52)	3452	56.27 (29.94)	7095	46.96 (26.24)	2425	–	–
Instances										
1	73.68 (26.42)	849	–	–	–	–	–	–	53.86 (26.61)	785
2	61.98 (27.15)	661	–	–	–	–	–	–	57.30 (27.28)	882
3	60.73 (26.56)	657	–	–	–	–	–	–	56.13 (26.63)	850
4	71.93 (25.84)	597	–	–	–	–	–	–	54.16 (26.53)	827
Total	67.45 (27.13)	2764	–	–	–	–	–	–	55.42 (26.80)	3344

Note. *N*s for reaction time and confidence differ due to exclusions of outlier reaction time values (~2% of data).

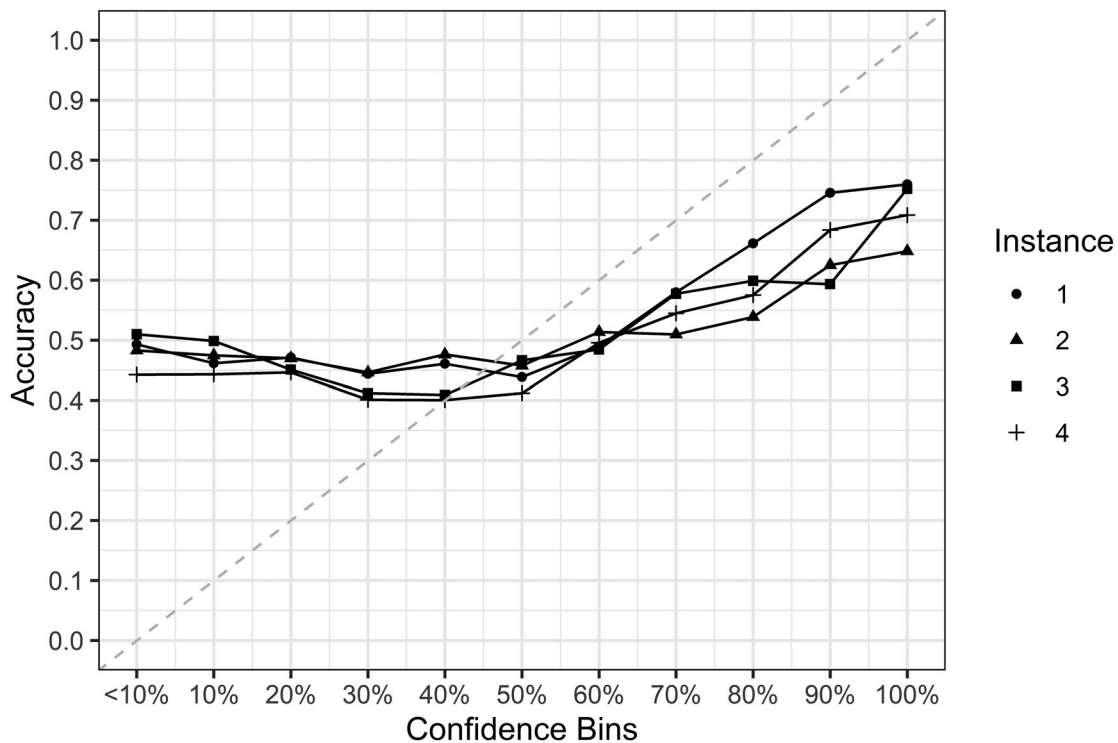
Table 5. Proportions of recognition decisions across conditions and instances in Experiment 2.

Decision	Task							
	Event				Instances			
	1	2	3	4	1	2	3	4
Old trials								
Hit	.68	.65	.63	.60	.52	.43	.44	.42
Miss	.32	.35	.37	.40	–	–	–	–
Misattribution	–	–	–	–	.48	.57	.56	.58
New trials								
Correct rejection	.74	.74	.75	.75	–	–	–	–
False alarm	.26	.26	.25	.25	–	–	–	–

participants needed to make source-attribution decisions, old items from the first and final instances were recognized faster than old items from the adjacent middle instances. Participants' confidence ratings were also higher for items from the boundary instances. Metacognitive measures further indicated that when participants were highly confident, their confidence better aligned with decision accuracy, particularly for items from the first instance. Further in line with the primacy effect, participants also more frequently judged items from the first instance as remembered, accurately recognized more old items from the first instance, and their decisions showed better source discriminability.

We used novel methodology combining paradigms that enabled us to measure reaction times in a task focused on recognition of items and assess serial position effects more broadly than in previous research (e.g.,

Dilevski et al., 2021a, who focused on discriminability). Further, we showed that ROC curves can be effectively constructed based on reaction time and confidence data to indicate differences in discriminability in terms of old and new items and discriminability in terms of source attribution accuracy across instances of repeated events. This approach is recommended in the recognition literature (e.g., Brady et al., 2023) and has clear potential for use in repeated event memory research. In our study, ROC curves indicated differences in discriminability consistent with the serial position effects. In future research, ROC curves can be used in investigations focused on the impact of deviation (i.e., unexpected changes occurring within an event) on recall, where it is frequently expected that deviation may increase instance memorability (e.g., Brubacher et al., 2011; Connolly et al., 2016). Next, collecting confidence statements associated with memory

**Figure 6.** Confidence-accuracy calibration curves between conditions and across instances in Experiment 2.

Note. Dashed reference line represents perfect calibration.

responses has potential in differentiating metacognitive abilities across instances of repeated events (see Roberts & Higham, 2002). In the present study, CA curves indicated better calibration for the boundary instances, but confidence-accuracy relationship can be similarly valuable in investigations of changes in metacognition as a consequence of instance deviation.

Our findings contribute to the growing evidence for primacy and recency effects in memory for repeated events, including the notions that: (i) primacy effects are typically stronger and more stable than recency effects (e.g., Dilevski et al., 2021a; Rubínová et al., 2022), and (ii) recency effects are sometimes absent (e.g., MacLean et al., 2018). Which mechanisms contribute to these effects? Rubínová et al. (2022) described that a combination of novelty (only the first instance), unique encoding context (e.g., Lohnas et al., 2015), and consequently reduced interference contribute to more effective source monitoring for items from the boundary instances (e.g., Johnson et al., 1993; Lindsay, 2008). Individuals likely form stronger memories of the first (and final) instances of repeated events that enable faster recognition of items and contribute to higher accuracy, discriminability, confidence, and more frequent judgments of remembering. On the other hand, items from the middle instances are likely involved in the process of script confirmation (e.g., Farrar & Boyer-Pennington, 1999) or generalisation and face increased interference due to high overlap of source cues, consequently limiting the effectiveness of source monitoring.

Our findings open new directions in repeated event memory research. In the repeated event memory literature, which is largely focused on applied questions of the ability of individuals to recall instances of repeated events (e.g., Guadagno et al., 2006; Woiwod & Connolly, 2017), the semantic (script-based) nature of remembering repeated events is typically highlighted (e.g., Deck & Paterson, 2021; Dilevski et al., 2021b; Kuebli & Fivush, 1994). On the other hand, the autobiographical literature conceptualizes repeated events as falling in between semantic and episodic memory (e.g., Bontkes et al., 2023; Renoult et al., 2012). Speculatively, our findings may indicate differences across instances of repeated events in the degree of where they fall within episodic and semantic memory. It is possible that memory for the first (and final) instances retain more episodic aspects (e.g., they are remembered in greater detail), while the middle instances may become more semantic (e.g., they are remembered in less detail).

In terms of applied implications, findings of the present study also show potential of using confidence statements accompanying memory-based decisions to improve inferences about report accuracy (e.g., Roberts & Higham, 2002). At high levels of confidence, items from first instances were associated with better confidence-accuracy calibration compared to items from other instances; in other words, statements of high confidence associated

with items from first instances may serve as more reliable indicators of accuracy. It should be noted that all CA calibration curves showed overconfidence at higher levels of confidence (i.e., participants' confidence ratings were higher than accuracy), although it is likely that the overall placement of the CA curves was influenced by the task. Specifically, in the present study, participants provided their statements following a forced-choice recognition decision where they did not have the option to withhold their response (or indicate "don't know"). It is possible that in recall paradigms that permit participants' regulation of memory output, the overall confidence-accuracy calibration would improve.

We would like to acknowledge limitations of the present research. First, we presented repeated event stimuli and collected memory responses within a single session. Can we expect that a repeated event presented within a single session would be encoded as a sequence of instances rather than as a single event? In an analysis of experiments using single-session presentation as well as presentation of each instance on a separate day, patterns of accuracy and consistency were similar across different procedures (Rubínová et al., 2022), although the examination of differences across procedures was not the focus in that study. Future research should examine how encoding of repeated events differs between single- and multiple-session presentation formats. Price et al. (2006) investigated the impact of spacing on recall of play sessions (four sessions in one day/ across four days) and found that spacing led to superior memory performance only when the delay to interview was short (1 d) but not when the delay was long (1 week). Danby et al. (2023) reported no difference for spacing in a sample of adults who watched sequences of videos, and they also noted that in their study, accuracy was at ceiling and error rates were at floor. Therefore, it is not clear if spacing would have an impact on our results, and we believe that more research is needed to understand spacing effects with the use of more complex stimuli (i.e., real events rather than videos). Relatedly, we only used one memory test with a short delay. Dilevski et al. (2021a) measured recognition performance three times with increasing delays (short, delay, 1, and 3 weeks), and found that differences in discriminability diminished at the long delay (3 weeks). Therefore, it remains to be established how stable effects found in this study would be at longer delays.

The second limitation directly follows: we used video stimuli, therefore generalisability to real-life repeated events may be limited. In an analysis of accuracy and consistency of recall across studies with different complexity of stimuli from wordlists to interactive events, Rubínová et al. (2022) reported consistent patterns of effects across stimuli. However, as Bontkes et al. (2023) highlighted, real-life repeated events occur under much more variable conditions including variability in place where instances of the repeated event occur. Three of the four stimuli sets we used in the present study did not change place across

instances (one did). Therefore, in terms of similarity of place, it is likely that we used predominantly one type of repeated event stimuli, and our findings may have limited generalisability to repeated events with greater variability of place (see Bontkes et al., 2023).

Third, this study was conducted online, which may have impacted participants' metacognitive judgments, particularly in terms of setting criterion. The presence of an authority figure, as may be the case with an in-person study, might shift participants towards stricter criterion (see relevant research on the impact of warnings on reducing the size of the misinformation effect; Wyler & Oswald, 2016), and we should be wary of this impact when generalising the current findings.

Fourth, we selected new items from a pool of plausible alternatives, but we did not test for lexical and semantic characteristics, such as word frequency or semantic similarity. However, we believe that the high variability of our stimuli (i.e., the use of four sets with counterbalanced order of instances and therefore items) provided a strong test of the primacy and recency effects. Finally, our samples included predominantly White women. Further research should focus on more diverse populations to see if our findings would generalize.

In conclusion, the present research found primacy (and recency) effects in long-term recall of instances of repeated events across a wide range of measures, indicating that first (and partially final) instances of repeated events are encoded with more attributes that uniquely link details to the first (final) instance. When the recognition test required engagement of source monitoring, old items from the boundary instances were recognized faster and with higher confidence compared to their adjacent middle instances. Primacy, but not recency effects, were found in memory judgments (items from first instances were more frequently "remembered"), accuracy, discriminability, and confidence-accuracy calibration. Our findings bring further evidence that individuals form stronger memories of first (and final) experiences of repeated events.

Author contribution statement

Eva Rubínová: Conceptualisation, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualisation, writing – original draft, writing – review and editing. **Heather L. Price:** Conceptualisation, funding acquisition, writing – review and editing.

Author Note

Preregistration including an analysis plan, data, and scripts can be accessed on the Open Science Framework (OSF) at <https://osf.io/hr2n3> (Experiment 1) and <https://osf.io/vxqcs> (Experiment 2). Partial results of this research were presented as a poster at the 63rd Annual Meeting of the Psychonomic Society, Boston, MA in November 2022. E. R. was

at Thompson Rivers University when the study was conducted. The majority of analyses and the entire writing was done at the University of Aberdeen.

Notes

1. Dilevski et al. (2023) did not find evidence for primacy and recency effects when asking participants which domestic violence scenario they remembered best, possibly because more than half of their participants selected one scenario that took place at a different location than other scenarios. Scenarios were presented randomly, so it is possible that the lack of serial position effects was at least partially due to the stimuli confound – if participants remembered the scenario because of the different location, this location distinctiveness may have driven participants' selection.
2. Our estimates of means were based on typical patterns found in correct recall (e.g., Rubínová et al., 2022), and within ranges of reaction times from recognition studies (Ratcliff & Murdock, 1976). The specified pattern of means across Instances 1–4 was 0.73, 0.68, 0.68, 0.71 for the event condition, and 0.93, 0.88, 0.88, 0.91 for the instances condition, with a common $SD = 0.27$. This pattern reflects the estimated differences across instances and conditions, although we erroneously flipped the effects – the values indicated *longer* reaction times for the boundary instances (an error due to correspondence to typical accuracy patterns), while we expected *shorter* reaction times for the boundary instances. The corrected pattern: 0.63, 0.68, 0.68, 0.65 for the event condition, and 0.83, 0.88, 0.88, 0.85 for the instances condition maintains estimated effect sizes and reaches identical estimates of required sample size.
3. We computed a sensitivity analysis to evaluate statistical power in the experiment. We used actual sample size, obtained mean reaction times across instances and conditions (following data exclusions and transformations), selected the largest SD for the common SD , calculated the intraclass correlation (package performance, Lüdtke et al., 2021) as an estimate of within-subjects correlation (Killip et al., 2004), $ICC = .682$, and set $\alpha = .03$. The sensitivity analysis indicated 100% power for detecting the effect of condition, 99% power for detecting the effect of instance, and 97% power for detecting the interaction. Therefore, our experiment had sufficient statistical power to detect the expected effects.
4. Two new items were removed from analyses of Set 3 because identical items were presented in two instances. One old item was removed from Set 4 because of a typo.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by a Natural Sciences and Engineering Research Council Discovery Grant to H.L.P. [NCERC grant: 2017-04986] and the Canada Research Chairs Program (H.L.P.).

References

- Ahn, W. K., Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 391–412. <https://doi.org/10.1037/0278-7393.18.2.391>

- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79(2), 97–123. <https://doi.org/10.1037/h0033773>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal statistical society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, 12, 675558. <https://doi.org/10.3389/fpsyg.2021.675558>
- Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, 13(1), 80–97. [https://doi.org/10.1016/S0022-5371\(74\)80033-2](https://doi.org/10.1016/S0022-5371(74)80033-2)
- Bontkes, O. R., Palombo, D. J., & Rubínová, E. (2023). Similarity Impacts Where Repeated Events Fall on the Semantic-Episodic Continuum [Unpublished manuscript [link to preprint will be supplied]].
- Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, 30(2), 421–449. <https://doi.org/10.3758/s13423-022-02179-w>
- Brown, G. D. A., & Lewandowsky, S. (2005). Serial recall and presentation schedule: A micro-analysis of local distinctiveness. *Memory (Hove, England)*, 13(3–4), 283–292. <https://doi.org/10.1080/09658210344000251>
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576. <https://doi.org/10.1037/0033-295X.114.3.539>
- Brown, G. D. A., Vousden, J. I., & McCormack, T. (2009). Memory retrieval as temporal discrimination. *Journal of Memory and Language*, 60(1), 194–208. <https://doi.org/10.1016/j.jml.2008.09.003>
- Brown, N. R. (2016). Transition theory: A minimalist perspective on the organization of autobiographical memory. *Journal of Applied Research in Memory and Cognition*, 5(2), 128–134. <https://doi.org/10.1016/j.jarmac.2016.03.005>
- Brubacher, S. P., Glisic, U., Roberts, K. P., & Powell, M. (2011). Children's ability to recall unique aspects of one occurrence of a repeated event. *Applied Cognitive Psychology*, 25(3), 351–358. <https://doi.org/10.1002/acp.1696>
- Cimbalo, R. S., Nowak, B. I., & Stringfield, C. (1978). Isolation effect: Overall list facilitation and debilitation in short-term memory. *The Journal of General Psychology*, 99(2), 251–256. <https://doi.org/10.1080/00221309.1978.9710510>
- Connolly, D. A., Gordon, H. M., Woiwod, D. M., & Price, H. L. (2016). What children recall about a repeated event when one instance is different from the others. *Developmental Psychology*, 52(7), 1038–1051. <https://doi.org/10.1037/dev0000137>
- Cooper, H. (2018). *Reporting quantitative research in psychology: How to meet APA style journal article reporting standards*. American Psychological Association.
- Danby, M. C., Brubacher, S. P., Sharman, S. J., Powell, M. B., & Roberts, K. P. (2017). Children's reasoning about which episode of a repeated event is best remembered. *Applied Cognitive Psychology*, 31(1), 99–108. <https://doi.org/10.1002/acp.3306>
- Danby, M. C., Sharman, S. J., van Golde, C., Paterson, H. M., & Watkins, R. (2023). The effects of episode spacing on adult's reports of a repeated event. *Memory (Hove, England)*, 31(6), 879–889. <https://doi.org/10.1080/09658211.2023.2198265>
- Deck, S. L., & Paterson, H. M. (2021). Adults also have difficulty recalling one instance of a repeated event. *Applied Cognitive Psychology*, 35(1), 286–292. <https://doi.org/10.1002/acp.3736>
- Dilevski, N., Paterson, H. M., & van Golde, C. (2020). Investigating the effect of emotional stress on adult memory for single and repeated events. *Psychology, Public Policy, and Law*, 26(4), 425–441. <https://doi.org/10.1037/law0000248>
- Dilevski, N., Paterson, H. M., & van Golde, C. (2021a). Adult memory for instances of a repeated emotionally stressful event: Does retention interval matter? *Memory (Hove, England)*, 29(1), 98–116. <https://doi.org/10.1080/09658211.2020.1860227>
- Dilevski, N., Paterson, H. M., & van Golde, C. (2022). Adult memory for instances of emotionally stressful and non-stressful repeated events. *Memory (Hove, England)*, 30(5), 621–635. <https://doi.org/10.1080/09658211.2022.2038630>
- Dilevski, N., Paterson, H. M., & van Golde, C. (2023). 'Tell me about the time you remember the best': The effect of a remember best prompt on adults' reports of a repeated emotionally stressful event. *Psychology, Crime & Law*, 29(4), 437–463. <https://doi.org/10.1080/1068316X.2022.2027945>
- Dilevski, N., Paterson, H. M., Walker, S. A., & van Golde, C. (2021b). Adult memory for specific instances of a repeated event: A preliminary review. *Psychiatry, Psychology and Law*, 28(5), 711–732. <https://doi.org/10.1080/13218719.2020.1837031>
- Doolen, A. C., & Radvansky, G. A. (2021). A novel study: Long-lasting event memory. *Memory (Hove, England)*, 29(8), 963–982. <https://doi.org/10.1080/09658211.2021.1953079>
- Ebbinghaus, H. (1884/1964). *Memory: A contribution to experimental psychology* (H. Ruger & C. E. Bussenius, Trans.). Dover Publications, Inc.
- Estes, W. K. (1985). Memory for temporal information. In J. A. Michon, & J. L. Jackson (Eds.), *Time, mind, and behavior* (pp. 151–168). Springer.
- Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organisation: A model of the von restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 224–240. <https://doi.org/10.1037/0278-7393.21.1.224>
- Farrar, M. J., & Boyer-Pennington, M. E. (1999). Remembering specific episodes of a scripted event. *Journal of Experimental Child Psychology*, 73(4), 266–288. <https://doi.org/10.1006/jecp.1999.2507>
- Farrar, M. J., & Goodman, G. S. (1990). Developmental differences in the relation between scripts and episodic memory: Do they exist? In R. Fivush, & J. Hudson (Eds.), *Knowing and remembering in young children* (pp. 30–64). Cambridge University Press.
- Farrar, M. J., & Goodman, G. S. (1992). Developmental changes in event memory. *Child Development*, 63(1), 173–187. doi:10.2307/1130911
- Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software*, 87, 1–27. <https://doi.org/10.18637/jss.v087.i09>
- Guadagno, B. L., Powell, M. B., & Right, R. (2006). Police officers' and legal professionals' perceptions regarding how children are, and should be, questioned about repeated abuse. *Psychiatry, Psychology and Law*, 13(2), 251–260. <https://doi.org/10.1375/pplt.13.2.251>
- Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 13(6), 644–655. [https://doi.org/10.1016/S0022-5371\(74\)80052-6](https://doi.org/10.1016/S0022-5371(74)80052-6)
- Henson, R. N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, 36(2), 73–137. <https://doi.org/10.1006/cogp.1998.0685>
- James, W. (1901). *Principles of psychology* (Vol. 1). Macmillan.
- 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>
- Killip, S. (2004). What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *The Annals of Family Medicine*, 2(3), 204–208. <http://doi.org/10.1370/afm.141>

- Kuebli, J., & Fivush, R. (1994). Children's representation and recall of event alternatives. *Journal of Experimental Child Psychology*, 58(1), 25–45. <https://doi.org/10.1006/jecp.1994.1024>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lakens, D., & Caldwell, A. R. (2021). Simulation-Based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, 4, 1–14.
- Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, 16(4), 395–418. [https://doi.org/10.1016/S0022-5371\(77\)80036-4](https://doi.org/10.1016/S0022-5371(77)80036-4)
- Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, 7(3), 149–169. <https://doi.org/10.1037/0278-7393.7.3.149>
- Lindsay, D. S. (2008). Source monitoring. In H. L. Roediger (Ed.), *Cognitive psychology of memory: Vol. 2. J. Byrne (Ed.), learning and memory: A comprehensive reference* (pp. 325–348). Elsevier.
- Lindsay, D. S. (2014). Memory source monitoring applied. In T. J. Perfect, & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 59–75). Sage.
- Loftus, E. F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*, 11(2), 114–120. <https://doi.org/10.3758/BF03213465>
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, 122(2), 337–363. <https://doi.org/10.1037/a0039036>
- Lüdtke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>
- MacLean, C. L., Coburn, P. I., Chong, K., & Connolly, D. L. (2018). Breaking script: Deviations and postevent information in adult memory for a repeated event. *Applied Cognitive Psychology*, 32(4), 474–486. <https://doi.org/10.1002/acp.3421>
- McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of experimental psychology: Learning, Memory, and Cognition*, 25(3), 563–582. <https://doi.org/10.1037/0278-7393.25.3.563>
- Morís Fernández, L., & Vadillo, M. A. (2020). Flexibility in reaction time analysis: Many roads to a false positive? *Royal Society Open Science*, 7(2), 190831. <https://doi.org/10.1098/rsos.190831>
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5), 482–488. <https://doi.org/10.1037/h0045106>
- Powell, M. B., & Thomson, D. M. (1997). Contrasting memory for temporal-source and memory for content in children's discrimination of repeated events. *Applied Cognitive Psychology*, 11(4), 339–360
- Powell, M. B., Thomson, D. M., & Ceci, S. J. (2003). Children's memory of recurring events: Is the first event always the best remembered? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 17, 127–146. <https://doi.org/10.1002/acp.864>
- Price, H. L., Connolly, D. A., & Gordon, H. M. (2006). Children's memory for complex autobiographical events: Does spacing of repeated instances matter? *Memory (Hove, England)*, 14(8), 977–989. <https://doi.org/10.1080/09658210601009005>
- Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, 17, 133–140. <https://doi.org/10.1016/j.cobeha.2017.08.006>
- Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83(3), 190–214. <https://doi.org/10.1037/0033-295X.83.3.190>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Renoult, L., Davidson, P. S., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences*, 16(11), 550–558. <https://doi.org/10.1016/j.tics.2012.09.003>
- Revelle, W. (2023). psych: Procedures for Psychological, Psychometric, and Personality Research. R package version 2.3.3. <https://CRAN.R-project.org/package=psych>.
- Roberts, K. P., Brubacher, S. P., Drohan-Jennings, D., Glisic, U., Powell, M. B., & Friedman, W. J. (2015). Developmental differences in the ability to provide temporal information about repeated events. *Applied Cognitive Psychology*, 29(3), 407–417. <https://doi.org/10.1002/acp.3118>
- Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied*, 8(1), 33–43. <https://doi.org/10.1037/1076-898X.8.1.33>
- Robinson, J. A. (1992). First experience memories: Contexts and functions in personal histories. In M. A. Conway, D. C. Rubin, H. Spinnler, & W. A. Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory* (pp. 223–239). Springer.
- Rothkopf, E. Z. (1971). Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behavior*, 10(6), 608–613. [https://doi.org/10.1016/S0022-5371\(71\)80066-X](https://doi.org/10.1016/S0022-5371(71)80066-X)
- Rubínová, E., Blank, H., Koppel, J., Dufková, E., & Ost, J. (2022). Repeated recall of repeated events: Accuracy and consistency. *Journal of Applied Research in Memory and Cognition*, 11, 229–244. <https://psycnet.apa.org/doi/10.1016/j.jarmac.2021.09.003>
- Rubínová, E., Blank, H., Koppel, J., & Ost, J. (2021). Schema and deviation effects in remembering repeated unfamiliar stories. *British Journal of Psychology*, 112(1), 180–206. <https://doi.org/10.1111/bjop.12449>
- Rubínová, E., Blank, H., Ost, J., & Fitzgerald, R. J. (2020). Structured word-lists as a model of basic schemata: Deviations from content and order in a repeated event paradigm. *Memory (Hove, England)*, 28(3), 309–322. <https://doi.org/10.1080/09658211.2020.1712421>
- Rubínová, E., & Kontogianni, F. (2023). Sources and destinations of misattributions in recall of instances of repeated events. *Memory & Cognition*, 51(1), 188–202. <https://doi.org/10.3758/s13421-022-01300-7>
- Rubínová, E., & Price, H. L. (2022a). *The cost of source monitoring in the recognition of instances of repeated events*. <https://doi.org/10.17605/OSF.IO/HR2N3>.
- Rubínová, E., & Price, H. L. (2022b). *The cost of source monitoring in recognition of instances of repeated events (Experiment 2)*. <https://doi.org/10.17605/OSF.IO/VXQCS>.
- Rubínová, E., Price, H. L., & Brubacher, S. P. (2023). *Prior Knowledge and the Recall of Single Events and Instances of Repeated Events: A Registered Report*. Retrieved from <https://osf.io/xrkct>.
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Schank, R. C. (1999). *Dynamic memory revisited*. Cambridge University Press.
- Sharman, S. J., Danby, M. C., & Christopoulos, L. (2022). Mental context reinstatement improves adults' reports of additional details from two instances of a repeated event. *Memory (Hove, England)*, 30(8), 988–999. <https://doi.org/10.1080/09658211.2022.2068610>
- Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin*, 124(3), 423–442. <https://doi.org/10.1037/0033-2909.124.3.423>
- Slackman, E., & Nelson, K. (1984). Acquisition of an unfamiliar script in story form by young children. *Child Development*, 55(2), 329–340. <https://doi.org/10.2307/1129946>

- Teigen, K. H., Böhm, G., Bruckmüller, S., Hegarty, P., & Luminet, O. (2017). Long live the king! Beginnings loom larger than endings of past and recurrent events. *Cognition*, 163, 26–41. <https://doi.org/10.1016/j.cognition.2017.02.013>
- Thomsen, D. K., & Berntsen, D. (2005). The end point effect in autobiographical memory: More than a calendar is needed. *Memory (Hove, England)*, 13(8), 846–861. <https://doi.org/10.1080/09658210444000449>
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology / Psychologie canadienne*, 26(1), 1–12. <https://doi.org/10.1037/h0080017>
- Tulving, E. (1989). Remembering and knowing the past. *American Scientist*, 77(4), 361–367. <https://www.jstor.org/stable/27855835>
- Underwood, B. J., & Freund, J. S. (1969). Further studies on conceptual similarity in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 8(1), 30–35. [https://doi.org/10.1016/S0022-5371\(69\)80007-1](https://doi.org/10.1016/S0022-5371(69)80007-1)
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S, Fourth edition*. Springer. <http://www.stats.ox.ac.uk/pub/MASS4/>
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25(9), 1115–1125. <https://doi.org/10.1177/01461672992512005>
- Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 77(1), 1–15. <https://doi.org/10.1037/h0028569>
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), 1–20. <http://www.jstatsoft.org/v21/i12>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>
- Wickham, H. (2022). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.0. <https://CRAN.R-project.org/package=stringr>
- Woiwod, D. M., & Connolly, D. A. (2017). Continuous child sexual abuse: Balancing defendants' rights and victims' capabilities to particularize individual acts of repeated abuse. *Criminal Justice Review*, 42(2), 206–225. <https://doi.org/10.1177/0734016817704700>
- Wyler, H., & Oswald, M. E. (2016). Why misinformation is reported: Evidence from a warning and a source-monitoring task. *Memory (Hove, England)*, 24(10), 1419–1434. DOI: [10.1080/09658211.2015.1117641](https://doi.org/10.1080/09658211.2015.1117641)