Repeated Recall of Repeated Events: Accuracy and Consistency

Eva Rubínová[a,b], Hartmut Blank[a], Jonathan Koppel[a], Eliška Dufková[c], James Ost[a]

[a]University of Portsmouth, Department of Psychology, King Henry Building, King Henry I Street, PO1 2DY, Portsmouth, Hampshire, United Kingdom

[b]Thompson Rivers University, Department of Psychology, 805 TRU Way, Kamloops, British Columbia, V2C 0C8, Canada

[c]Masaryk University in Brno, Department of Psychology, Faculty of Social Studies, Joštova 10, 602 00, Brno, Czech Republic

Author Note

Correspondence concerning this article should be addressed to Eva Rubínová, Department of Psychology, Thompson Rivers University, 805 TRU Way, Kamloops, British Columbia, V2C 0C8, Canada.

E-mail: erubinova@tru.ca

Abstract

In both casual conversations and interview settings, people may be required to provide details of instances that were similar to other experiences. When this happens repeatedly, consistency across reports is often taken as a proxy for credibility. However, processes of schema formation and interference due to similarity make recall and accurate source attribution of details to specific instances challenging. We investigated the accuracy and consistency of recall in these contexts in a re-analysis of five studies. Confusions of details were widespread (1) across instances—participants frequently attributed the origin of details to incorrect instances, but also (2) across repeated retrieval attempts—participants frequently changed parts of their reports. There was, however, a clear pattern of primacy and recency effects: recall of the first and final instances was more accurate and consistent than recall of the middle instances. We discuss potential mechanisms underlying these effects as well as their practical implications.

Keywords. Repeated events, repeated recall, internal intrusions, source monitoring, recall consistency, recall accuracy.

General Audience Summary

People may be asked to recall instances of events that were similar to other experiences, sometimes on several occasions. Importantly, consistent recall of specific experiences is difficult, as the similarity of each experience to one another makes it challenging to distinguish between them. We examined accuracy and consistency of recall of repeated events in a re-analysis of data from five studies. Repeated events consisted of four instances (interactive sessions, stories, or word-lists), and participants recalled each instance on several occasions with increasing delay.

Recall was generally a mix of details that happened during a specific instance and details that happened during other instances. However, participants were more accurate in their recall of the first and final instances, and more confused in their recall of the middle instances. For example, participants would frequently include details that occurred during Instance 3 in their report of Instance 2. Moreover, such confusions occurred not only across instances, but also across retrieval attempts: participants were more consistent in their recall of Instances 1 and 4 than in their recall of Instances 2 and 3. These primacy and recency effects for accuracy and consistency were stable across delay.

We discuss these findings in relation to processes that are engaged when a person tries to decide in which instance (i.e., source) specific details occurred. Source decisions are relatively straightforward regarding details of the first instance: its novelty as the establishing instance confers unique detail characteristics that guide source decisions. The experience of the instance that concluded the repeated event may also provide unique detail characteristics for this instance. Such source monitoring is, however, more difficult, and consequently erroneous for details of the middle instances, which lack unique characteristics. The result is low consistency, which should be considered in settings where repeated reports of repeated events are required.

Repeated Recall of Repeated Events: Accuracy and Consistency

In casual conversations about shared experiences or in investigative interviews, people may be required to recall details of several instances of a repeated event (e.g., Kelloway et al., 2003; Stark, 2012). Shopping, work, or lectures are examples of everyday repeated events (see Barsalou, 1988; Gioia & Poole, 1984; Neisser, 1988; Renoult et al., 2012); domestic violence or stalking are examples of repeated criminal offences (Sheridan et al., 2003; Stark, 2012); and industrial accidents are examples of incidents that often occur in the course of repeated events (Kelloway et al., 2003; MacLean et al., 2013). When people provide memory reports of single events on multiple occasions, credibility is often judged based on consistency (i.e., differences between reports are interpreted as indicators of low credibility; Brewer & Hupfeld, 2004; Cohen, 2001; Fisher et al., 2009; Granhag et al., 2005). Importantly, recall of instances that are similar to one another involves specific challenges to consistency, due to the consequences of repeated experiences on memory: If remembering (instances of) repeated events is difficult, doing so consistently over time is an even tougher challenge.

Memory for Instances of a Repeated Event

Instances of a repeated event share an underlying structure (e.g., language lessons involve an ordered set of typical exercises) but have limited overlap of details (e.g., the content of the exercises differs) and vary in contextual aspects (e.g., the presence/absence of other people). Schemata that emerge after repeated experience facilitate memory reconstruction of shared aspects of instances (e.g., Bartlett, 1932; Danby et al., 2019; Minsky, 1974; Rumelhart et al., 1986; van Kesteren et al., 2013; Schank, 1999), while accurate attribution of details that vary across instances depends on unique characteristics that may provide source links (Johnson et al., 1993; Lindsay, 2008; 2014). In case of repeated events, such links may lack specificity and face substantial interference caused by the overlap of shared aspects across instances (Postman, 1971). The result is confusion: in an attempt to recall details from a specific instance, people may recall a mix of details from this and other instances (i.e., internal intrusions; e.g., Woiwod et al., 2019). Some instances, however, seem to be less prone to these confusion errors.

Specifically, there are several reasons for expecting better memory for the *boundary* (i.e., first and last) instances. First, transition theory (N. R. Brown, 2016) holds that the beginnings and endings of prolonged or repeated events serve as landmarks that help maintain temporal organization within autobiographical memory (see also Robinson, 1992). This basic idea can be extended to the recall of repeated events occurring over shorter periods, and similarly leads to the expectation that the first and final instances possess unique qualities as reference points. Moreover, when compared to instances in between, memories for the boundary instances receive less interference from their neighbours. Both of these effects would enhance source memory for the boundary instances.

Additionally, memory for the first instance may be facilitated via involuntary reminding due to repeatedly experiencing similar instances (Hintzman, 2011) and during schema formation (Farrar & Boyer-Pennington, 1999; Farrar & Goodman, 1992). The resulting primacy and recency effects have been widely documented as a pattern of higher proportions of correct details and fewer detail confusions in the recall of the first and final instances of repeated events (Connolly et al., 2016; Dilevski et al., 2020a, 2020b; MacLean, et al., 2018; Powell & Thomson, 1997; Powell et al., 2003; Roberts et al., 2015; Rubínová et al., 2020). How do these patterns translate into consistency?

Consistency in Recall of Instances of a Repeated Event

Consistency reflects the correspondence between details provided at multiple retrieval attempts. Omission, reminiscence (i.e., recall of previously unreported information), and contradictory source attribution all cause inconsistency (Engelhard, et al., 2008; Odinot et al., 2013; Kaasa et al., 2011; Krinsley et al., 2003). From an applied perspective, contradictions in reports would likely decrease credibility judgments (e.g., Fisher, et al., 2009). When compared to single events, the potential for contradiction in recall of instances of repeated events is much greater due to source confusions that may occur not only across instances but also interviews (e.g., a set of details may be initially accurately attributed to the instance in which they occurred but later attributed to another instance).

Research investigating consistency within repeated events is sparse and researchers typically focus on recall of only one target instance (e.g., Price et al., 2016; Theunissen et al., 2017). To our knowledge, Connolly and Price (2013), who interviewed a victim of five bank robberies, are the only exception. Their findings were in line with primacy and recency effects: confusion of details across instances was the leading cause of inconsistency and was highest in the middle instances; however, the absence of ground truth in Connolly and Price's study precluded evaluation of accuracy.

Current Research

We investigated patterns of accuracy and consistency in repeated recall of instances of repeated events. In this secondary data analysis, we examined data from five studies in which our primary aim was to investigate effects of deviations on recall (we collapsed data from all conditions for the present purposes; see Method for further details). These studies varied in the complexity of stimuli: participants recalled interactive visits in Study 1, stories in Study 2, and categorized word-lists in Study 3, 4, and 5. In all studies, participants experienced or viewed four instances of the repeated event, and then were asked to recall all instances on a number of occasions with increasing delay. Across all studies, we assessed recall of specific details that were unique to instances. In previous analyses, the main recall measure reflected completeness (i.e., the proportion of accurately recalled details out of all presented details), and repeated retrieval attempts were examined only to see any temporal changes in recall.

The current study examined these data with a novel focus on *output accuracy* (i.e., the proportion of accurately recalled details out of all recalled details; Koriat et al., 2000) and consistency (i.e., the correspondence across pairs of recall attempts), which are most relevant in applied settings. Because Studies 2, 4 and 5 are already published (Rubínová et al., 2020, 2021) and the method of Study 3 was almost identical to Study 4, we only report a minimum of necessary information on the methods of these studies but provide a detailed description of Study 1.

We examined the patterns of overall accuracy, accuracy of consistent recall, and consistency. For consistency, we examined the patterns of consistent recall, contradictory attributions, omissions, and reminiscences. Given that primacy and recency effects are typically found in analyses of correct responses (e.g., MacLean et al., 2018), we expected to find these effects for accuracy, consistent recall, and, in the reverse, for contradictory attributions. We additionally explored recall patterns of omissions and reminiscences. Finally, we asked how much overlap is there in repeated recall of specific instances? To address this question, we took the subset of consistently recalled details and contradictory attributions (i.e., details that were recalled repeatedly). Given that omissions would be expected due to forgetting (and reminiscences are rare), a ratio of consistently recalled details and contradictory attributions would reflect the stability of source memory. In line with the expected primacy and recency effects, this ratio should indicate higher stability of source memory for the boundary than for the middle instances.

Study 1

Method

In Study 1, our aim was to create stimuli that would resemble real-life repeated events, in order to increase external validity relative to other laboratory-based repeated event studies (e.g., MacLean et al., 2018; Rubínová et al., 2020, 2021). Inspired by developmental repeated-event studies where children are typically interviewed about instances in which they directly participated (e.g., play sessions; Farrar & Boyer-Pennington, 1999), we created a series of marketing-themed visits that would be engaging for our target sample of adults. At each of four visits that occurred within a two-week interval, participants experienced the same three structured activities during which they interacted with, inspected, and evaluated products (for a full list of details, see Appendix A). Participants' recall from two recall sessions (one to two weeks after the final visit a month later) was assessed for accuracy and consistency.

Design

This study used a 4 (visit: 1/2/3/4) × 2 (recall session: first/second) within-subjects design. There were two additional between-subjects factors: 2 (content: typical/deviation) × 2 (order: typical/deviation) for experimental manipulations of content of one activity (i.e., we changed its purpose) and order of two activities in the final visit; because these manipulations are not of primary interest in this paper and because they had little effect on recall, they are not reported here (see Rubínová, 2020a). Moreover, the patterns of accuracy and consistency examined in participants from the non-deviation conditions were similar to those reported in this manuscript; see Figure SM5 in Online Supplemental Materials.

Participants

We analysed data from a study that primarily focused on deviation effects on recall of instances of a repeated event; therefore, we did not conduct an a-priori power calculation specifically for the accuracy or consistency analyses. The study was advertised as an investigation into people's memory for a series of marketing-themed visits. The advert mentioned that participants would be required to attend four marketing-themed visits, and that there would be an interview about their experiences one to two-weeks after the final visit.

Participants were informed that the purpose of the study was to learn how well can people remember details of a series of instances that are similar to each other.

A total of 191 participants were recruited. Eight participants did not complete all the visits, data from three participants were excluded due to errors during the administration of the visits, and 57 participants did not complete the second recall phase. The final sample therefore consisted of 124 participants (24 males, 89 females, 11 not indicated) aged between 18 and 57 years (*M* = 22.57, *SD* = 6.92, 11 participants did not indicate age). Participants were recruited from an undergraduate participant pool and from a university noticeboard, reported having normal or corrected-to-normal vision and hearing and English language fluency. The award for participation was 1 study credit and £5, or £15. The study was approved by the Science Faculty Ethics Committee (SFEC 2018-014).

Procedure

Visits. Before commencing the study, participants were informed that their task each time would be to inspect and evaluate a series of products, and that there would be an interview during which they would be asked what they could remember from each of the visits. They were also informed that there was a follow-up online survey that would be sent a month after the interview, in which they would be asked again what they remembered from the sessions and complete a few additional questions.

Individual visits were separated by delays: the average delay was 1.81 days (*SD* = 0.97) between Visits 1 and 2, 2.08 days (*SD* = 1.09) between Visits 2 and 3, and 1.81 days (*SD* = 0.99) between Visits 3 and 4. At each visit, participants experienced three activities: Game, Products, and Device. Table A in Appendix A shows four sets of stimuli including specific details that were used during the visits. The sets were counterbalanced across visits (e.g., during each visit, approximately one quarter of participants was administered the red, green, blue, or purple game, respectively). The visits were administered in a lab with a central room used for the Products, an adjacent room used for the Game, and another adjacent room used for the Device activity (i.e., participants changed rooms for each activity).

During the Game activity, participants played a demo of a story-telling game that involved rolling dice with pictures (story elements) on each side and then creating a story from these pictures. At each visit, the experimenter (E. R.) introduced the game version and played first. Participants played next and then completed a brief evaluation sheet, in which they wrote down the pictures they rolled and provided ratings for quality and enjoyment of the game. The game version and participants' story images constituted four specific details from this activity. The view of the table during play was video recorded (with participants' consent) to aid recall coding.

During the Products activity, participants evaluated the packaging design (logo and graphics) of three products that were placed on a table. The product category and specific products constituted four specific details from this activity (see Appendix A).

Device, which was the last activity, had three phases. Participants first read an introduction to the task from a laptop screen, then performed three tasks with the device according to displayed instructions, and finally completed an evaluation form. The device and details of the three tasks constituted four specific details from this activity (see Appendix A).

First Recall Session: Interview. One to two weeks after Visit 4 ($M = 9.51$ days, $SD = 2.37$), participants returned for an interview. The scripted interviews were administered in a different room by one of seven trained interviewers (58 interviews were administered by E. R. and 66 by one of the other six interviewers). The interviews had two phases. Participants were first asked to complete a written account of each visit using an A3 sheet with four windows (one for each visit). Participants were asked to write a complete report of each visit including as many details as possible without guessing. This format was intended to provide participants with a visual overview of all the visits (similar to the timeline technique; Hope, et al., 2019;

Hope, Mullis, & Gabbert, 2013). To provide participants with an additional free-recall prompt and an opportunity to reflect upon their written accounts, we asked them to provide a verbal account of each visit with the use of their written protocols. Verbal reports of each visit were followed by an open prompt ("Is there anything else you can remember about this visit?"), and there was one final prompt at the end of the interview ("Is there anything else you would like to report?"). Interviews were audio-recorded.

Written and verbal phases of the recall interview were coded separately and then combined. The combined measure included all information that was omitted during the verbal phase but mentioned in the written phase, and all information that was additionally recalled or changed in the verbal phase (there were few differences between the written and verbal reports; for a comparison, see Rubínová, 2020a).

Second Recall Session: Online Form. Four weeks after the interview, participants were emailed an online recall form that consisted of four blank pages and instructions similar to those used during the interview. The average delay between the first and the second recall was 35.70 days (*SD* = 6.70).

Coding

Details.[1] In each visit, there were 12 specific details: (1) the game version or the color of the box; (2 – 4) three pictures on the dice; (5) the product category; (6 – 8) three branded products; (9) the device; and (10 – 12) three specifications of the device task (see Appendix A). For one detail of each activity, our coding accommodated coarse-grained responses that sufficiently differentiated between the visits (i.e., the recall of the color of the box for game, product category for products, and any description that could be identified as the device for devices). For the remaining tree details of each activity, fine-grained recall was necessary for

[1] We also coded recall of activities (i.e., game, products, device); however, this variable mainly reflected the schematic representation of instances (see Rubínová, 2020a). Given our focus on recall accuracy and consistency of fine details, this variable was not of interest in this study.

coding recall as a specific detail (e.g., “pictures” was insufficient for pictures on the dice; “grey deodorant” was insufficient for branded products; and “highlighting” was insufficient for the device task when “underline”, “cross out”, and “circle around” were the specific tasks). In other words, our coding scheme accommodated recall at various levels of specificity. For a more detailed description and coding manuals, see Online Supplemental Materials.

Details were coded as *accurate* (i.e., experienced details attributed to a visit in which they occurred), *internal intrusions* (i.e., experienced details attributed to a visit in which they did not occur), and *new details* (i.e., details that did not occur during any of the visits). We encountered a few cases where participants reported the same detail at multiple instances (e.g., “Fantasia game” was recalled as occurring in Visit 1 and in Visit 2), and also a few cases where participants reported a detail that contained aspects of several details that originated at different visits (e.g., “We played the Actions version that had a green box”, when in fact Actions was blue, and Voyages was green). We coded these occurrences as details with “conflicting source” because the source was either undetermined or mixed. These details (3% of overall recall) represented confused source information; therefore, we collapsed them with internal intrusions. Finally, we coded a few occurrences of vague descriptions of Products, where reported information was insufficient to indicate a specific detail (e.g., “chocolate” or “deodorant for men”). In these cases, recall would be coded for the category of Products (either as “correct” or “internal intrusion” depending on attribution), and we added a “vague” code for one of the specific products, but we treated “vague” codes (2% of overall recall) as not recalled details in statistical analyses.

Coding was a two-step semi-automatic procedure. Briefly, rich verbal reports were first reduced and validated according to a coding manual (e.g., “white labelling pen” was validated as “white marker”). Validated data were then automatically coded against a reference sheet; the exception were details pertaining to pictures of the dice game, which were coded manually.

Data were entered, validated, and (where applicable) manually coded by E. D., and her scores were used for further consistency and accuracy coding and for all statistical analyses. To obtain estimates of inter-rater reliability, two subsets of the data were independently coded by trained research assistants (Device and Product details from written reports of 114 participants) and E. R. (Device and Game details from written reports of 126 participants). Comparing each of these subsets to E. D.'s coding resulted in high agreement (Cohen's kappa ranged between 0.90 and 0.97).

Accuracy. Accuracy reflected the correspondence of recalled details to details encountered during the visits. Accurate recall included details that were experienced and attributed to the visit in which they occurred. Inaccurate recall included internal intrusions and new details. Note that this coding scheme reflected "narrowly defined" accuracy as opposed to "broadly defined" accuracy (Price et al., 2016), where only new details would be considered inaccurate and internal intrusions would be considered accurate (see Price et al.; Woiwod et al., 2019). We report results pertaining to broadly defined accuracy in the Online Supplemental Materials.

Consistency. Consistency was coded based on the correspondence of recalled details in the first and second recall sessions. Consistent recall included details that were recalled and attributed to the same visit at both recall attempts. Inconsistent recall included: (i) contradictory attributions, (ii) omissions of previously recalled details, and (iii) reminiscences of previously not reported details.

Measures

Accuracy. Recalled details were coded as accurate or inaccurate (accuracy values of 1 and 0, respectively).

Consistency. Details consistently recalled and attributed in both recall sessions were coded as consistent (value of 1) and all other details (i.e., contradictory attributions, omissions, and reminiscences) were coded as inconsistent (value of 0).

Consistent Recall: Accuracy. We assessed accuracy of recall in the subset of details that were consistently recalled across sessions.

Contradictory attributions. Details that were repeatedly recalled across both recall sessions, but each time attributed to a different visit were coded as contradictory attributions (value of 1); all other details (i.e., consistent details, omissions, and reminiscences) were coded as not contradictory attributions (value of 0).

Omissions. Details that were omitted in the second recall session were coded as omissions (value of 1); all other details (i.e., consistent details, contradictory attributions, and reminiscences) were coded as non-omissions (value of 0).

Reminiscences. Details that were reported only in the second recall session were coded as reminiscences (value of 1); all other details (i.e., consistent details, contradictory attributions, and omissions) were coded as non-reminiscences (value of 0).

Stability Ratio. To assess the stability of source memory for each instance, we took the subset of details that were recalled at both recall sessions. In this subset, we divided the proportion of consistently recalled details by the proportion of contradictory attributions to compute a ratio.

Statistical Analyses

Data were analyzed in R (R Core Team; 2016) with random intercepts generalized mixed models using the `lme4` (Bates et al., 2015) and `lmerTest` packages (Kuznetsova et al.; 2017). Our design was fully crossed—all participants recalled details of four visits at two occasions. To account for dependencies in the data and avoid pseudo replication, each model included random intercept effects for participant identity and specific detail (Finch et al., 2014).

We built models for six outcome variables: (1) overall accuracy; (2) consistency; (3) accuracy of consistent recall, (4) contradictory attributions, (5) omissions, and (6) reminiscences. For the ratio of recall stability, we computed bootstrapped 95% Confidence Intervals [*CI*s] using package `boot` (Canty & Ripley, 2021), and we evaluated differences between the ratios based on non-overlapping *CI*s (Tryon, 2001).

All dependent variables were binomial, and all fixed and random effects were categorical. The fixed effects of recall session and visit (i.e., instance) were coded with successive difference contrasts (Schad et al., 2020) from the `MASS` package (Venables & Ripley, 2002). The first contrast for visit compared Visits 1 and 2, the second contrast compared Visits 2 and 3, and the third contrast compared Visits 3 and 4. The first and the third contrasts were of utmost interest in this study because they assessed local primacy and recency effects as relative differences between the boundary instances and their adjacent instances.[2] The contrast for recall session compared the two sessions.

Regression coefficients from all models were exponentiated, so the resulting coefficients indicated the odds ratios (*OR*s) of one outcome over another between groups defined by respective contrasts (e.g., consistent vs inconsistent recall in Visits 1 and 2). Any significant interactions indicated moderating effects of recall session on accuracy in a given visit contrast. We present the *OR*s along with 95% *CI*s in brackets to indicate the range of their plausible values (Cumming, 2012, 2014). To aid interpretation of *OR* values below 1 we also present the inverted *OR*s (OR_i). We used Ferguson's (2009) guidance for interpreting

[2] Interested readers may use the code and data provided in Online Supplemental Materials and examine other contrasts that may be of interest.

effect sizes: *OR* < 3.00 as small, *OR* = [3.00, 4.00] as moderate, and *OR* ≥ 4.00 as strong effects.

We used packages `ggplot2` (Wickham, 2016) and `gridExtra` (Auguie, 2017) for all visualizations. Data, coding functions, and analysis scripts are available in Online Supplemental Materials.

Results

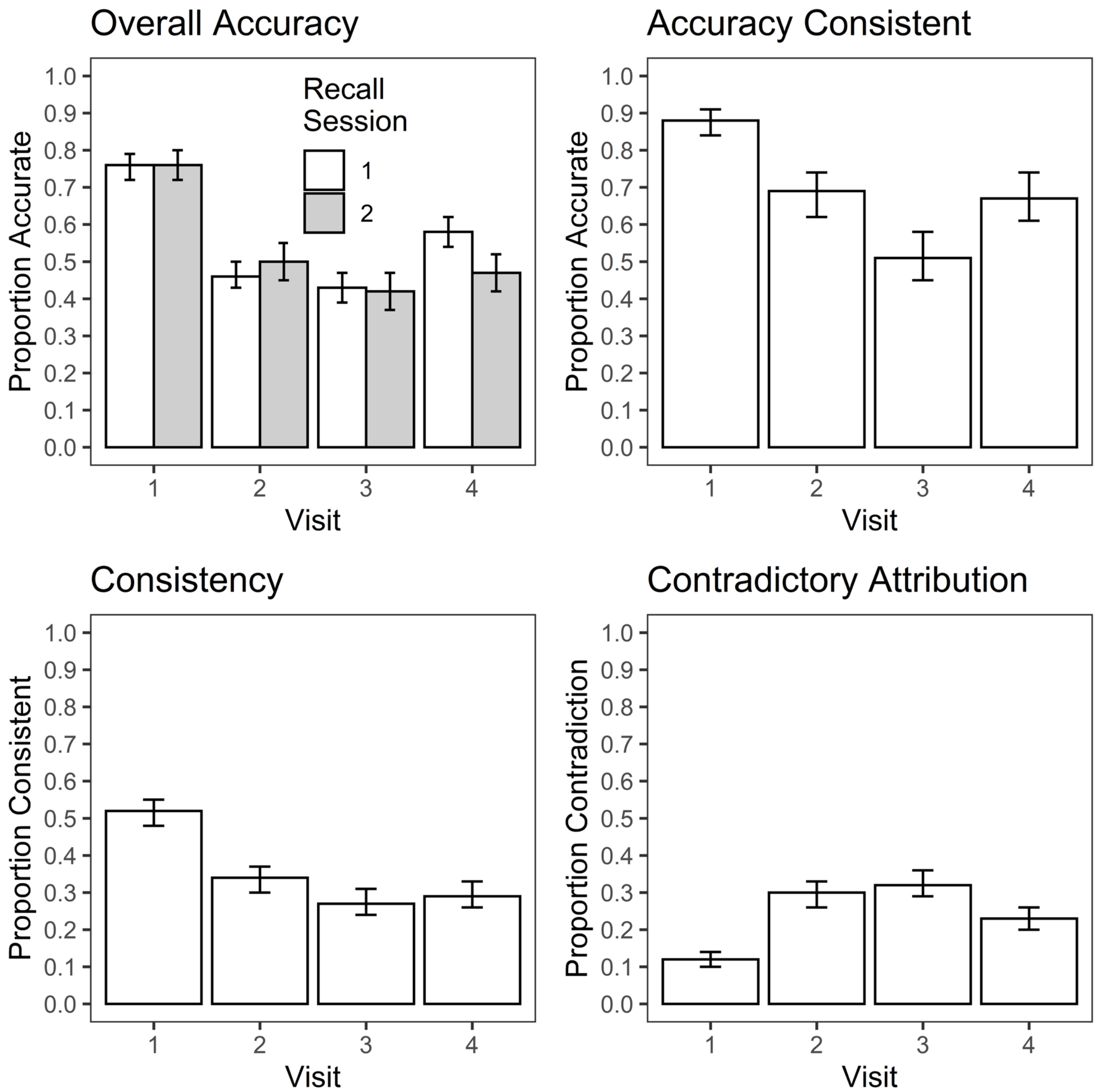
Due to our focus on recall patterns that would be most relevant for applied settings, the results reported in this section are output-bound (Koriat et al., 2000). In other words, all measures were coded and analysed based on details that were recalled at the two recall sessions. To get a perspective on general recall performance across visits and recall sessions, see Figure SM1 in the Online Supplemental Materials.

Overall Accuracy

Figure 1 shows the patterns of overall accurate recall across visits in the two recall sessions. The results confirmed the visible strong primacy effect, OR_i = 4.00 [3.28, 4.88], z = 13.72, $p < .001$. The recency effect was also significant though small, *OR* = 1.64 [1.34, 2.00], z = 4.87, $p < .001$. The effect of recall session was significant, with recall being slightly less accurate at Session 2, OR_i = 1.20 [1.04, 1.39], z = 2.52, p = .012. This effect was mainly driven by the absence of a recency effect in the second recall session, which is reflected in the significant Visit 4-3 × Recall Session interaction, OR_i = 1.68 [1.13, 2.49], z = 2.57, p = .010. Note that the decrease in accuracy across recall sessions indicated an increase in source attribution errors (i.e., source memory) because accuracy (as a proportion of accurate details among all recalled details) did not reflect forgetting of details. There were no further significant effects (full statistics are reported in Table SM2 in Online Supplemental Materials). When accuracy was broadly defined (i.e., internal intrusions were considered accurate), the proportions of accuracy across all visits and both recall sessions were uniformly high (i.e., between 0.94 and 0.97; see Table SM3 in Online Supplemental Materials).

Figure 1

Overall accuracy, accuracy of consistent recall, consistency, and contradictory attributions across visits and recall sessions in Study 1



Note. Error bars represent bootstrapped 95% *CI*s of the proportions.

Consistency

In general, consistency reflects the correspondence of recall across recall sessions. Figure 1 shows the patterns of overall consistency across visits, and the analysis confirmed a small-to-moderate primacy effect, OR_i = 2.65 [2.06, 3.40], z = 7.61, p < .001but the recency effect was not significant, OR = 1.22 [0.93, 1.60], z = 1.46, p = .145 (full statistics are reported in Table SM2 in Online Supplemental Materials). There are, however, different types of consistent and inconsistent recall. Table 1 lists (1) the proportions of consistent recall based on detail source (i.e., details consistently correctly attributed to instances in which they occurred, details consistently attributed to incorrect instances, and consistently attributed new details), and (2) the proportions of inconsistent recall based on detail source (in case of omissions and reminiscences), respectively source attribution changes (in case of contradictory attributions).

There are several patterns worth noting in Table 1. First, there was considerable variability in the accuracy of consistently reported details: consistent recall for Visit 1 was highly accurate, but for Visit 3, almost half of consistent recall composed of internal intrusions. Second, contradictory attributions (i.e., source attribution changes) most often occurred from or back to accurate attribution (i.e., participants most frequently misattributed details that were originally attributed to correct instances or the other way round); changes of source attribution that were incorrect at both recall sessions (i.e., details from Visit 1 were first recalled in Visit 2 and then in Visit 3) more frequently occurred in the middle than at the boundary instances. Third, omissions show differences in types of details across instances: correctly attributed details composed higher proportions of omissions from the boundary instances, whereas internal intrusions composed higher proportions of omissions from the middle instances. Finally, reminiscences and new details were infrequent. In the next sections, we present analyses of overall consistency, accuracy of consistent recall, and overall contradictory attributions, omissions, and reminiscences (i.e., the fine differences in different types of inconsistent recall are beyond the scope of this study).

Table 1

Proportions of types of consistent and inconsistent recall in Study 1

Recall	Visit 1	Visit 2	Visit 3	Visit 4	$N_{Details}$
Consistent					
Correct	0.45	0.23	0.14	0.20	735
Intrusion	0.05	0.10	0.13	0.09	264
New	0.02	0.01	< 0.01	< 0.01	22
Inconsistent					
Contradictory Attribution					
From Correct	0.07	0.12	0.14	0.13	328
To Correct	0.02	0.08	0.10	0.04	177
Within Intrusions	0.02	0.09	0.07	0.05	164
New	0.01	0.01	0.01	0.01	25
Omission					
Correct	0.21	0.10	0.13	0.24	481
Intrusion	0.09	0.20	0.18	0.17	463
New	0.03	0.03	0.03	0.04	96
Reminiscence					
Correct	0.02	0.01	0.03	0.01	45
Intrusion	0.01	0.01	0.03	0.02	53
New	< 0.01	0.01	< 0.01	0.01	23

Note. Intrusion = internal intrusion.

Accuracy of Consistent Recall. Figure 1 shows the patterns of accuracy of consistent recall across visits. The analyses confirmed the strong primacy, $OR_i = 4.06$ [2.51, 6.55], $z = 5.73$, $p < .001$ and small-to-moderate recency effects, $OR = 2.72$ [1.62, 4.56], $z = 3.79$, $p < .001$ (full statistics are reported in Table SM2 in Online Supplemental Materials). The recency effect was, however, only a local effect dependent on the lower accuracy for Visit 3 (i.e., the absolute proportions of accurate recall were similar for Visits 2 and 4). Broadly defined accuracy of consistent recall was uniformly high across visits (i.e., between 0.97 and 0.99; see Table SM3 in Online Supplemental Materials).

Contradictory Attributions, Omissions, and Reminiscences. The analysis of overall contradictory attributions visible in Figure 1 confirmed the (reverse) moderate primacy, *OR* = 3.58 [2.66, 4.81], $z = 8.42$, $p < .001$, and small recency effects, $OR_i = 1.82$ [1.40, 2.37], $z = 4.45$, $p < .001$ (full statistics are reported in Table SM2 in Online Supplemental Materials). Overall omissions indicated a uniform pattern across Visits 1, 2, and 3; the odds of omissions were significantly higher for Visit 4 than for Visit 3 but the effect was small, *OR* = 1.83 [1.37, 2.44], $z = 4.11$, $p < .001$ (full statistics are reported in Table SM4 in Online Supplemental Materials). Reminiscences were rare (see Table 1) and uniform (see Table SM4 in Online Supplemental Materials).

Stability Ratio

The ratio of stability of instance source memory was computed as a proportion of consistently recalled details and contradictory attributions. A perfect stability would occur if details that were repeatedly recalled were attributed to the same instance at both recall sessions. A stability ratio of 1 would indicate that the number of details attributed to the same instance at both recall sessions was equal to the number of details that were attributed to different instance(s) at the second recall session (i.e., 50% stability). Values above 1 would indicate higher stability of source memory; values below 1 would indicate lower stability of source memory.

For Visit 1, the value of the stability ratio, 4.33 [3.20, 5.33], indicated that there were four details consistently reported for one contradictory attribution (i.e., 81% stability). This stability ratio is higher than for any other visit: Visit 2, 1.14 [0.93, 1.33]; Visit 3, 0.84 [0.69, 1.00]; and Visit 4, 1.29 [1.01, 1.55]. For Visits 2 and 4, the stability ratios indicated that there were almost as many consistently reported details as contradictory attributions (i.e., 53 – 56% stability). For Visit 3, the stability ratio indicated that contradictory attributions were in fact slightly more frequent than consistently reported details, indicating relatively low stability of source memory (46%). The primacy and recency effects were significant, as indicated by non-overlapping confidence intervals between Visits 1 and 2 (primacy) and Visits 3 and 4 (recency, although this effect was again only local due to the low stability ratio for Visit 3).

Discussion

In Study 1, we investigated accuracy and consistency in repeated recall of four instances of a repeated event involving complex interactions. For overall accuracy, we found primacy and recency effects; the recency effect was, however, reduced at the second recall session. We found similar patterns of accuracy within consistent recall. When accuracy was broadly defined (Price et al., 2016), recall was almost perfectly accurate, confirming that participants reported details that they experienced within the repeated event but were confused about when exactly these details occurred.

The novel findings of this study pertain to patterns of consistency. For overall consistency, we found a primacy (i.e., recall of Visit 1 was most consistent) but no recency effect. It seems that in the final instance, consistency was reduced primarily by omissions and not by contradictory attributions, a notion that was supported by the recall stability ratios. The overall patterns suggested that source memory was indeed stronger for the boundary instances, and that the primacy effect was much larger than the recency effect (for a similar finding, see Connolly & Price, 2013). Reminiscences were low, with small differences across instances.

Before we further discuss potential mechanisms underlying these effects, we examined data from another four studies that used the repeated event paradigm (albeit with simpler materials and slightly different procedures), to see if the patterns would generalize.

Studies 2 – 5

We re-examined data from four studies and coded them for accuracy and consistency. Materials used in these studies were much simpler than in Study 1 and lacked any interactive aspects: participants were presented versions of an unfamiliar story in Study 2 and categorized word-lists in Studies 3, 4, and 5. In addition, there were several methodological differences (for details, see Method). Therefore, finding similar results regarding accuracy and consistency in these studies would increase the reliability and generalizability of findings from Study 1.

Method

In this section, we briefly summarize the methods of Studies 2 – 5 (see Table 2). For parsimony, we only focus on aspects that were of interest in the present study (i.e., we do not describe any additional features). The design of all studies was the same as Study 1, including content and order manipulations in the final instance. In Study 2, the content deviation was implemented as a change in a part of a Story, and in Studies 2 – 5, we presented a new word-category. The order deviation was implemented as a changed temporal order of themes in a story (Study 2) or word-categories (Studies 3 – 5). The deviations had little effect on recall in Study 3 (see Rubínová 2020b). The content deviation details were well remembered in Study 4 but there was no significant effect in Study 5 (Rubínová et al., 2020); in Study 2, the content deviation details were frequently distorted to fit with the other stories (Rubínová et al., 2021). The order deviation effect had a generally disruptive though small effect on recall (Rubínová et al., 2002, 2021). An examination of accuracy and consistency in a subset of data from participants from the non-deviation conditions revealed similar patterns to those reported in this article (see Figure SM5 in Online Supplemental Materials). Therefore, data from all conditions were collapsed for the purpose of the present analysis.

All studies were advertised as memory studies. Participants were informed that they would be presented with four stories (Study 2), respectively word-lists (Studies 3 – 5), and that they should pay attention to the stimuli because they would recall them later. Participants completed the first delayed recall task in the lab and received links to follow-up online answer forms after specific delays (see Table 2). Participants were told that the purpose of these studies was to learn about how repeated exposure to similar events affects memory for specific instances of those events. Participants were not informed about the purpose of the online answer forms but were told that their completion would take 10 to 15 minutes.

There were three main procedural differences from Study 1 (see Table 2). First, all instances of the repeated events were presented during a single session. Second, in Studies 2, 4, and 5, participants were asked to recall each instance before the next instance was presented.[3] Third, in Studies 2, 3, and 4, delayed recall of all instances was measured in four intervals.

Table 2

Details of methods in Studies 2 – 5

Study	Reference	*N*	Stimuli	Details	Recall Before Next Instance	Delayed Recall
2	Rubínová et al. (2021)	148	Versions of an unfamiliar story	11	Yes	10 min, 1 D, 1 W, 1 M
3	Rubínová (2020b)	101	Categorized word-lists	9	No	10 min, 1 D, 1 W, 1 M
4	Rubínová et al. (2020), Study 1	80	Categorized word-lists	9	Yes	10 min, 1 D, 1 W, 1 M
5	Rubínová et al. (2020), Study 2	96	Categorized word-lists	9	Yes	10 min, 1 D

Note. D = day. W = week. M = month.

Study 2

[3] In Studies 2, 4, and 5, participants additionally recalled each instance one minute after encoding. In line with the testing effect (e.g., Roediger & Karpicke, 2006), this isolated recall (absent in Studies 1 and 3) was used to facilitate participants' memory of each instance (a comparison of performance indicated that recall was indeed higher in Study 4 than in Study 3; Rubínová, 2020b). In addition, the isolated recall limited interference and in turn enabled accurate detail attribution—in Studies 4 and 5, the immediate isolated recall was almost perfectly accurate (see data from Rubínová et al., 2020) and in Study 2, the emerging schema facilitated recall across instances (i.e., recall of Instance 2 was higher than recall of Instance 1; Rubínová et al., 2021). However, this immediate recall did not entirely prevent source confusion on subsequent recall tasks. Therefore, any confusion of details across instances (and related primacy and recency effects) occurred after a delay when participants were asked to recall all instances, and where their memory likely relied on the reconstruction of instances based on new schemata and detail attribution required systematic source monitoring. For these reasons, we excluded the immediate isolated recall from the current study.

Materials. We adapted a passage describing a Korean wedding ceremony from Ahn, Mooney, and Brewer (1992). The passage was simplified, and four versions were created that shared a common structure in terms of sequences of actions but differed in 11 specific details (e.g., characters, names, or dates). The resulting stimuli were videos showing a sequence of illustrations accompanied by a narrative (see Online Supplemental Materials accompanying Rubínová et al., 2021).

Procedure. Participants viewed a video of Story 1 twice and, after a 1-minute filler task, recalled Story 1 (i.e., rehearsal/isolated recall before the next instance, which is not included in our current analysis; see Footnote 3). After a further 2-minute filler task, they proceeded to viewing Story 2. This procedure was repeated until participants recalled Story 4. A 10-minute filler task followed, after which participants recalled all four stories (recall session 1). In this delayed recall phase, participants were presented with four pages that contained illustrations of the two main characters and a designation of the story (e.g., Story 1). Participants could switch between the pages during recall. The same recall procedure was then administered approximately one day, one week, and one month after presentation (recall sessions 2, 3, and 4, respectively).

Coding. Accuracy codes included accurate details, internal intrusions, and new details. Cases of details where source attribution was unresolved (i.e., one detail was attributed to multiple instances) were more frequent than in Study 1 (15% of overall recall), but because they indicated source attribution errors, we collapsed them with internal intrusions (as in Study 1). Finally, we encountered a few cases (0.4% of overall recall) where actions were attributed to characters who did not perform them. Because these distortions represent a different type of confusion than internal intrusions, we collapsed them with new details. Based on these basic codes, consistency was then determined exactly as in Study 1. Note that consistency was coded for consecutive recall sessions; specifically, contradictory attributions occurring between recall sessions 1 and 2 and between recall sessions 2 and 3 (and 3 and 4) were different contradictory attributions.

Studies 3 - 5

Materials and Procedure. Stimuli were word-lists consisting of nine words from three ordered word-categories. In Studies 3 and 4, participants were told that they would see words that a language student learned on four consecutive days. The words were presented one-by-one. Each list was presented on a different background colour, and was designated by a photograph of the student and the day of the week (e.g., words that the student learned on Monday). The photograph, list designation, and background colour were used as cues in recall sessions. In Study 5, there was no cover story and lists were designated ordinally (e.g., List 1). The procedure was the same as in Study 2, with two exceptions: (i) in Study 3, participants did not recall the word-lists before the next word-list was presented; and (ii) in Study 5, there were only two recall sessions (10 minutes and one day after presentation).

Coding. Accuracy codes included accurate details, internal intrusions, and new details. Cases of details with unresolved source were infrequent (2%, 2%, and 3% in overall recall in Studies 3, 4, and 5, respectively), and were collapsed with internal intrusions. Consistency coding was the same as Study 2.

Results

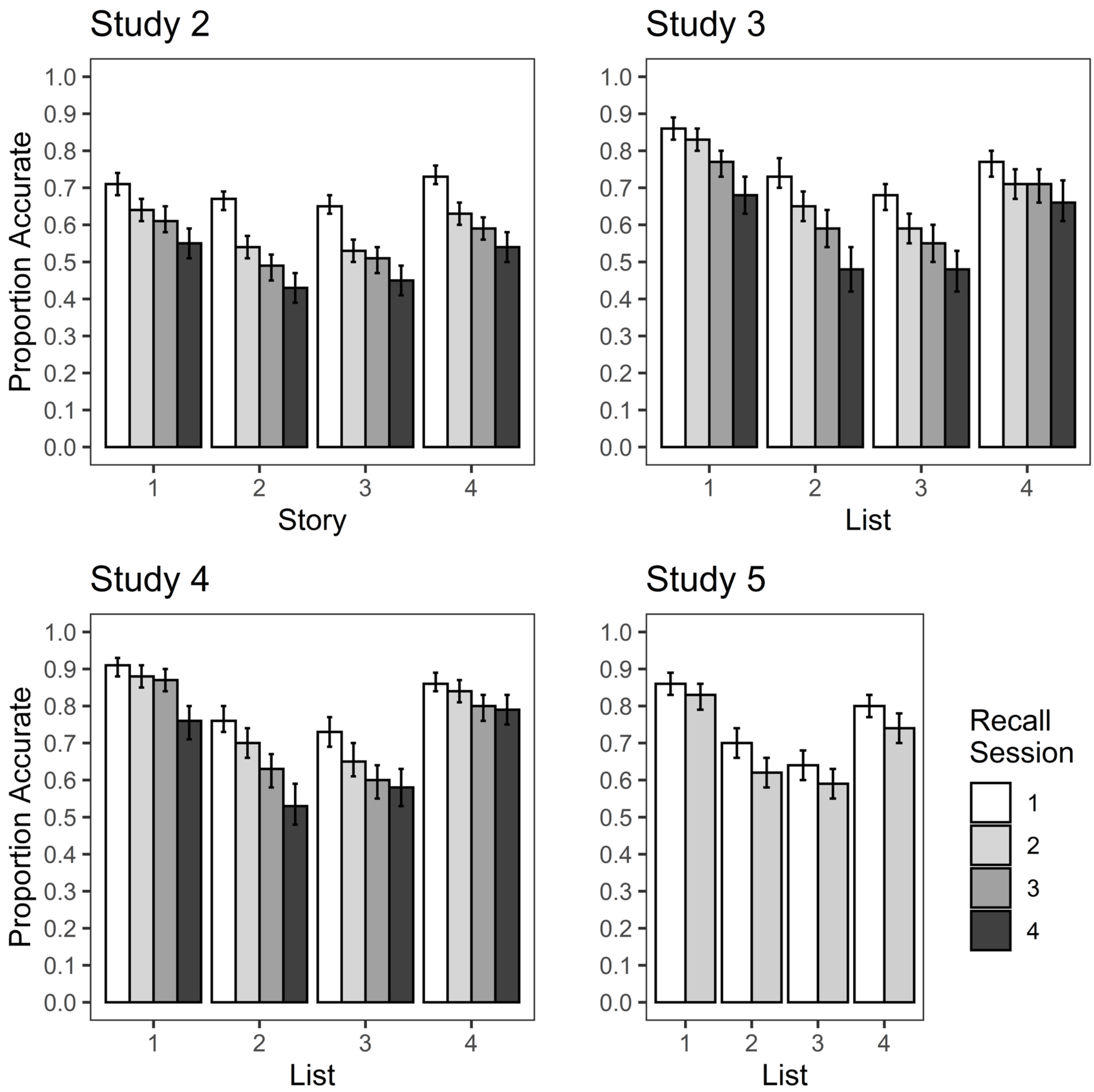
Overall Accuracy and Accuracy of Consistent Recall

Figure 2 shows the patterns of overall accuracy across instances and recall sessions. In all studies, the results indicated small-to-large primacy effects, $ORs_i = 1.62 – 4.23$, ps < .001, and small-to-medium recency effects, $ORs = 1.58 – 3.14$, ps < .001, and a small decrease in accuracy across recall sessions, $ORs_i = 1.43 – 1.74$, ps < .001 (note that the decrease in accuracy across sessions reflects incorrect source attribution and not forgetting). None of the interactions between visit and recall session were significant, indicating that the primacy and recency effects were stable across delay. The difference between the middle visits was hardly ever significant (and very small; full statistics are reported in Table SM5 in Online Supplemental Materials). Broadly defined accuracy across instances and sessions was high (between 0.85 and 1.00; see Table SM6 in Online Supplemental Materials).

The patterns and results for accuracy of consistent recall resembled those reported above. To avoid repetition, we present Figure SM2 and Tables SM7 (results of accuracy models) and SM6 (broadly defined accuracy) in Online Supplemental Materials.

Figure 2

Overall accuracy across instances and recall sessions in Studies 2 – 5



Note. Error bars represent bootstrapped 95% *CI*s of the proportions.

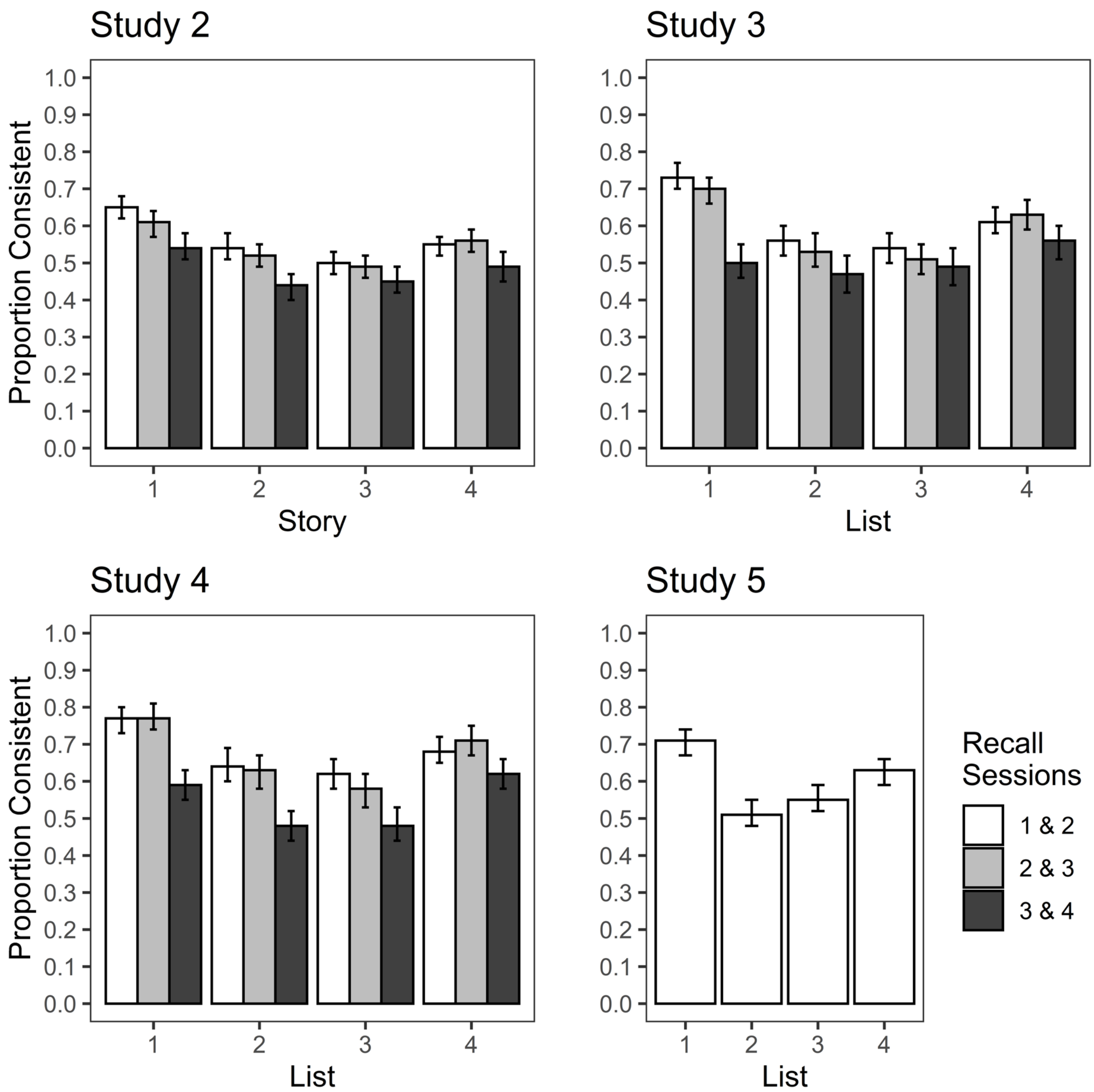
Consistency

Figure 3 shows consistency across instances and recall sessions. The results confirmed small-to-moderate primacy effects, ORs_i = 1.66 – 2.76, *ps* < .001, small recency effects, *ORs* = 1.27 – 1.83, *ps* < .001, and a small decrease in consistency across sessions, ORs_i = 1.37 – 1.56, *ps* < .001 (full statistics are reported in Table SM8 in Online Supplemental Materials;

note that the decrease in consistency across recall sessions is a result of contradictory attributions, omissions, and reminiscences).

Figure 3

Consistency across instances and recall sessions in Studies 2 – 5



Note. Error bars represent bootstrapped 95% *Ci*s of the proportions.

Contradictory Attributions, Omissions, and Reminiscences. The patterns of contradictory attributions were essentially mirror images of the patterns of consistency;

therefore, we report Figure SM3 in the Online Supplemental Materials. The analysis confirmed small-to-large primacy effects, *ORs* = 1.52 – 4.70, *ps* < .001, and small recency effects, ORs_i = 1.46 – 1.66, *ps* ≤ .007. Contradictory attributions were stable across delay; the effect of recall session was significant only in Study 4, where it indicated a small increase in the odds of contradictory attributions across delay, *OR* = 1.22 [1.08, 1.38], *z* = 3.24, *p* = .001 (full statistics are reported in Table SM9 in Online Supplemental Materials).

For omissions, there was a consistent small primacy effect indicating lower omissions for Instance 1 than Instance 2, *ORs* = 1.34 – 1.72, *ps* ≤ .001, and a small increase in omissions across sessions (indicating forgetting), *ORs* = 1.41 – 1.87, *ps* < .001. The patterns and statistical results for omissions are reported in Figure SM4 and Table SM10 in the Online Supplemental Materials. Reminiscences were infrequent and their patterns were uniform (for proportions and analysis results, see Tables SM11 and SM12 in Online Supplemental Materials).

Stability Ratio

Stability ratios across instances and recall sessions are reported in Table 3. These ratios are higher than in Study 1, likely due to methodological differences (specifically, in Study 1, the delay to first and second recall was longer, which should lower stability). In general, there are again moderate-to-large primacy and small-to-large recency effects indicating higher source memory stability in the first and the final instances.

Table 3

Ratios of instance memory stability in Studies 2 – 5

Study/ Recall Sessions	Instance 1	Instance 2	Instance 3	Instance 4
Study 2				
1&2	3.88 [3.23, 4.54]*	2.17 [1.85, 2.48]	1.71 [1.48, 1.92]	2.63 [2.25, 3.00]*
2&3	3.20 [2.61, 3.76]	2.40 [2.00, 2.77]	1.95 [1.64, 2.25]	2.91 [2.41, 3.37]*

3&4	2.90 [2.32, 3.42]*	1.61 [1.32, 1.88]	1.65 [1.37, 1.91]	2.25 [1.84, 2.64]
Study 3				
1&2	12.13 [7.67, 15.97]*	4.12 [2.96, 5.17]	3.54 [2.63, 4.34]	6.26 [4.37, 7.90]*
2&3	19.43 [9.33, 27.47]*	4.24 [2.98, 5.34]	3.33 [2.42, 4.17]	9.34 [5.54, 12.54]*
3&4	7.83 [4.54, 10.72]	3.94 [2.59, 5.12]	4.25 [2.75, 5.55]	5.98 [3.68, 7.85]
Study 4				
1&2	34.23 [7.25, 55.53]*	6.43 [4.39, 8.21]	4.87 [3.52, 6.08]	7.00 [5.07, 8.75]
2&3	26.19 [8.16, 40.92]*	5.29 [3.55, 6.76]	4.03 [3.02, 5.00]	8.56 [5.43, 11.24]*
3&4	13.91 [6.61, 20.00]*	3.00 [2.15, 3.77]	3.01 [2.17, 3.82]	6.52 [4.11, 8.50]*
Study 5				
1&2	19.25 [9.45, 27.19]*	3.52 [2.64, 4.29]	3.82 [2.91, 4.65]	6.22 [4.44, 7.74]

Note. Sessions = recall sessions. * = significant difference as indicated by non-overlapping *CI*s of adjacent instances.

Discussion

Our systematic re-analysis of studies that used the repeated event paradigm with massed presentation and stimuli as simple as categorized word-lists replicated the findings from Study 1. Specifically, we obtained primacy and recency effects for accuracy, accuracy of consistent recall, contradictory attributions, and recall stability, as well as the relatively uniform patterns of reminiscences and broadly defined accuracy. Regarding consistency, in addition to the primacy effect, Studies 2 – 5 also indicated a recency effect. In contrast with Study 1, where omissions were higher in Instance 4, the patterns in Studies 2 – 5 indicated slightly reduced omissions for Instance 1.

Internal Meta-Analysis

In this final section, we provide a brief meta-analytical overview of the primacy and recency effects found in Studies 1 – 5 for measures of overall accuracy, accuracy of consistent recall, consistency, and contradictory attributions.

Method

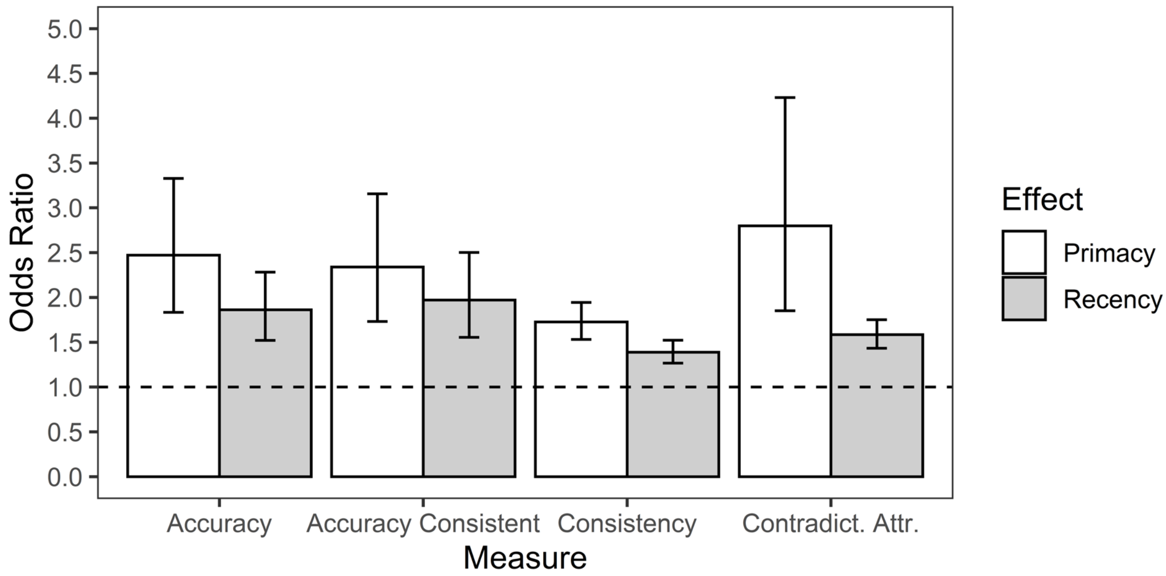
To prepare data for the internal meta-analyses of primacy and recency effects in Studies 1 – 5, we collapsed across participants and recall sessions in all studies. We then calculated proportions for: (1) overall accuracy as the sum of accurate details divided by the sum of inaccurate details; (2) accuracy of consistently reported details as the sum of accurate consistently reported details divided by the sum of inaccurate consistently reported details; (3) consistency as the sum of consistent details divided by the sum of inconsistent details; and (4) contradictory attributions as the sum of contradictory attributions divided by non-contradictory attributions. Using the package `metaphor` (Viechtbauer, 2010), we calculated random effects models for the primacy effects by comparing the outcomes between Instances 1 and 2, and for the recency effects by comparing the outcomes between Instances 3 and 4. The resulting *OR*s indicate the sizes of the effects.

Results

Figure 4 illustrates the resulting odds ratios for each effect. For overall accuracy, the primacy effect was small-to-moderate, 2.47 [1.83, 3.33], and the recency effect was small, 1.86 [1.52, 2.28]. For accuracy of consistent recall, the primacy effect was also small-to-moderate, 2.33 [1.73, 3.16], and the recency effect was small, 1.97 [1.56, 2.50]. For consistency, the primacy effect was small, 1.73 [1.53, 1.94], and the recency effect was small, 1.39 [1.27, 1.52]. Finally, for contradictory attributions, the primacy effect was small-to-moderate, 2.80 [1.85, 4.23], and the recency effect was small, 1.58 [1.43, 1.75].

Figure 4

Results of internal meta-analyses for selected measures in Studies 1 – 5



Note. Error bars represent 95% *CI*s. The dashed line indicates no effect. Contradict. Attr. = contradictory source attribution.

General Discussion

Our findings can be summarized as follows. For accuracy and consistency in repeated recall of repeated events, there were (stronger) primacy and (weaker) recency effects throughout. Although accuracy and consistency slightly decreased overall across recall sessions, these primacy and recency effects remained stable. Recall stability ratios indicated that source memory for the boundary instances was preserved better than for the middle instances. There were no consistent effects for omissions. Reminiscences were uniformly low, and broadly defined accuracy was uniformly high.

Serial Position Effects in Short- and Long-Term Recall

The patterns we observed in long-term recall, with recall delays sometimes exceeding a month, resemble typical serial position effects in short-term recall tasks (e.g., Bjork & Healy, 1974; G. D. A. Brown et al., 2007; Healy, 1974; Lewandowsky & G. D. A. Brown, 2005). In these tasks, participants recall items from a list, and accurate performance requires retained memory for order. Likewise, in our studies, participants recalled details of instances, and accurate performance required retained source memory. Cognitive models developed for short-term serial recall tasks explain primacy and recency as isolation effects—items at list boundaries have greater local distinctiveness (i.e., limited exposure to interfering items; G. D. A. Brown et al., 2009), and this explanation also holds for our long-term memory data. There are, however, also important differences.

In short-term serial recall, the primacy effect is typically smaller than the recency effect, the recency effect decreases with delay, and the primacy effect is explained mainly via rehearsal (G. D. A. Brown et al., 2007). By contrast, in our repeated event paradigms, rehearsal of the first instance was unlikely because of the spacing of instances in Study 1 and the immediate recall following each instance in Studies 2, 4, and 5. Our primacy effects more likely result from novelty and schema formation (e.g., Farrar & Boyer-Pennington, 1999; Hintzman, 2011) – specifically, the first instance becomes a reference (N. R. Brown, 2016) and acquires unique source attributes that protect it from source confusion (Lindsay, 2008).

Schema Formation and Source Monitoring in Repeated Events

Memory reconstruction of an event involves the activation of coarse-grained knowledge-based structures (i.e., schemata) along with the retrieval of fine-grained information (i.e., specific details). Retrieval of specific instances involves judgments about the inclusion of details into the final memory report, and attributes of memories generated as a product of "cognitive processes performed during a particular past event" provide the basis for these judgments (Lindsay, 2008, p. 328). Such source monitoring can be automatic if the event was relatively unique but may require systematic decision-making if the event substantially overlapped with other event(s) (Johnson et al., 1993; Lindsay, 2008). Importantly, effective source monitoring requires that attributes of memories from different sources are distinguishable (Johnson et al., 1993); repeated experiences, however, likely violate this assumption.

Repeated events start with an initial instance that does not require the creation of a new schema; for most events, existing schemata will be sufficient for understanding what happens and for later memory reconstruction (e.g., Bartlett, 1932). Nevertheless, for familiar events, the first experience constitutes a novel schema instantiation, and this novelty serves as a unique source attribute for details of the first instance. In addition, the encoding context of the first experience that establishes the repeated event (e.g., Lohnas et al., 2015; Kahana, 1996; Robinson, 1992), when retrieved, may also serve as a unique source attribute that limits misattributions.

The experience of the second instance is different: it is likely reflected upon as another event similar to the first instance, and this reflection sets off the generation of a new schema specific for this emerging repeated event (e.g., Farrar & Boyer-Pennington, 1999; Slackman & Nelson, 1984). The experiences of the third and fourth instances then reinforce the schema, and source attributes of these "other similar instances" may be insufficient for differentiation between instances. Consequently, in contrast with the first instance, more details of the following instances may be generally linked to the repeated event (schema) rather than to specific instances. The exception is the final instance: In cases where participants know that an instance is final (as in the present studies), this status as the concluding instance constitutes a unique encoding context and generates attributes that provide at least partial protection from interference. Along these lines, the recall patterns in our studies can be summed up by three simple assumptions illustrated in Table 4: The novelty of the first instance, and the unique encoding context and limited interference from similar neighbours promote correct source attribution for the boundary instances. The combined result of these mechanisms is a strong primacy and a weak recency effect.

Table 4

Mechanisms determining source memory for instance of repeated events

	Effect on Source Memory			
Mechanisms and Results	Instance 1	Instance 2	Instance 3	Instance 4
Mechanism				
Novelty	+	0	0	0
Unique Encoding Context	+	0	0	+
Interference	-	--	--	-
Result				
Combined	+	--	--	0
Boundary Relative to Middle	+++	0	0	++

Note. + = positive effect on result. - = negative effect on result. 0 = no effect.

Limitations

Several limitations of Studies 2 – 5 have already been discussed elsewhere (see Rubínová et al., 2020, 2021). A limitation specific to Study 1 relates to output regulation[4] (e.g., Koriat & Goldsmith, 1996). We only recognized responses that enabled differentiation of details, and for three details in the product category, this often required the recall of the brand (i.e., recall of the colour of the packaging but not the brand would not pass the specific detail threshold, and would be counted as an omission). In applied settings, though, more coarse-grained responses could be considered, and might result in higher accuracy and consistency estimates. Therefore, our approach could have resulted in higher omission rates than in interviewing settings, but we do not expect that the general pattern of accuracy and consistency across instances would change because patterns of omissions showed only small (if any) differences across instances. Importantly, patterns of contradictory attributions would remain the same.

A further limitation applies to all studies: Some real-life repeated events do not have a clear beginning (e.g., grooming or experiences that gradually escalate). In such cases, it is not

[4] In Studies 2 – 5, coding fine-grained responses was our only option because coarse-grained responses would be equivalent to describing the script of the stories or the schema of the word-lists and would therefore not be informative enough to differentiate between instances.

clear if we would observe a strong primacy effect. But even so, one instance could become a landmark and would therefore be associated with an effect similar to primacy.

Practical Implications

In recall of instances of repeated events, we saw that confusion of details occurs not only across instances, but also across retrieval attempts. In the middle instances in Study 1, disregarding omissions and reminiscences, only 50% of details of the same instance were shared in repeated recall, and the remaining 50% were transferred to or from other instances. Such low consistency results from the way memory for repeated events is organized, that is, as a schema that allows only limited discrimination between instances, producing errors in source monitoring.

Further, consistency and accuracy decreased across repeated retrieval attempts. Specifically, detail attributions often transitioned from initially accurate to subsequently inaccurate (Table 1). Correspondingly, we recommend that interviewers elicit an initial account of instances of repeated events and, wherever possible, limit repeated interviewing. Where repeated questioning is necessary, however, interviewers can maximize the likelihood of consistent and accurate recall by focusing on the first or the final instances. Our findings showed primacy (and, more weakly, recency) effects to be stable over time; therefore, we should expect superior memory particularly for the first instance even when statements are made after longer intervals, which is quite likely in applied contexts.

Finally, our findings call for more realistic expectations regarding source memory in applied contexts: Instead of dismissing inconsistent reports or regarding inconsistent interviewees as lacking credibility, it should be acknowledged that imperfect memory is not a sign of unreliability but rather a consequence of the hierarchical organization of memory for instances of repeated events. Completely abandoning the requirement of particularization in cases of repeated offences (see Woiwod & Connolly, 2017) might be premature, but it would certainly reflect the practical difficulty associated with repeated recall of repeated events. Future research could further explore more realistic credibility assessments that do not rely on exaggerated expectations regarding source memory for repeated events.

Author Contributions

Eva Rubínová developed the concept, designed the experiments, collected the data, conducted all statistical analyses, and drafted the manuscript. Hartmut Blank and James Ost provided critical feedback on the concept and design of the experiments. Eliška Dufková entered, developed coding functions, and coded all data from Experiment 1. All authors (except for the late James Ost) contributed to revisions of the manuscript.

Acknowledgements

Experiment 1 was supported by the Society for Applied Research in Memory and Cognition Student Caucus Research Grant, Experiment 2 was supported by the International Centre for Research in Forensic Psychology Travel Grant, and all experiments were supported by the University of Portsmouth postgraduate bursary to the first author. The authors would like to thank Jagoda Mizera, Matthew Seal, Keeley Stevens, Lucia Horváthová, Rafaela Brandao for their help with data collection, Tomáš Rubín for his help with data collection and coding functions, and Heather Price for her comments on an earlier draft.

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*, 391-412. https://doi.org/10.1037/0278-7393.18.2.391

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics.* R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 193-243). New York: Cambridge University Press.

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48. https://doi.org/10.18637/jss.v067.i01

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*, 80-97. https://doi.org/10.1016/S0022-5371(74)80033-2

Brewer, N., & Hupfeld, R. M. (2004). Effects of Testimonial Inconsistencies and Witness Group Identity on Mock-Juror Judgments 1. *Journal of Applied Social Psychology*, *34*, 493-513. https://doi.org/10.1111/j.1559-1816.2004.tb02558.x

Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. Psychological Review, *114*, 539-576.

Brown, G. D., Vousden, J. I., & McCormack, T. (2009). Memory retrieval as temporal discrimination. *Journal of Memory and Language*, *60*, 194-208.

Brown, N. R. (2016). Transition theory: A minimalist perspective on the organization of autobiographical memory. *Journal of Applied Research in Memory and Cognition, 5*, 128-134. https://doi.org/10.1016/j.jarmac.2016.03.005

Brown, G. D., & Lewandowsky, S. (2005). Serial recall and presentation schedule: A micro-analysis of local distinctiveness. *Memory*, *13*, 283-292. https://doi.org/10.1080/09658210344000251

Canty, A., & Ripley, B. D. (2021). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-26.

Cohen, J. (2001). Questions of credibility: omissions, discrepancies and errors of recall in the testimony of asylum seekers. *International Journal of Refugee Law*, *13*, 293-309.

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*, 1038-1051. https://doi.org/10.1037/dev0000137

Connolly, D. A., & Price, H. L. (2013). Repeated interviews about repeated trauma from the distant past: A study of report consistency. In *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 191-217). Springer, New York, NY.

Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Hove, UK: Routledge.

Cumming, G. (2014). The new statistics why and how. *Psychological Science*, *25*, 7-29. https://doi.org/10.1177/0956797613504966

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*, 99-108. https://doi.org/10.1002/acp.3306

Dilevski, N., Paterson, H. M., & van Golde, C. (2020a). Investigating the impact of stress on adult memory for single and repeated events. *Psychology, Public Policy, and Law*. https://doi.org/ 10.1037/law0000248

Dilevski, N., Paterson, H. M., & van Golde, C. (2020b). Adult memory for instances of a repeated emotionally stressful event: Does retention interval matter? *Memory*. https://doi.org/10.1080/09658211.2020.1860227

Engelhard, I. M., van den Hout, M. A., & McNally, R. J., 2008. Memory consistency for traumatic events in Dutch soldiers deployed to Iraq. *Memory*, *16*, 3–9. https://doi.org/10.1080/09658210701334022

Estes, W. K. (1985). Memory for temporal information. In J. A. Michon & J. L. Jackson (Eds.), *Time, mind, and behavior* (pp. 151-168). Berlin, GE: Springer.

Farrar, M. J., & Boyer-Pennington, M. E. (1999). Remembering specific episodes of a scripted event. *Journal of Experimental Child Psychology, 73*, 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). New York, NY: Cambridge University Press.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 532-538. https://doi.org/10.1037/a0015808

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. Boca Raton, FL: CRC Press.

Fisher, R.P., Brewer, N. & Mitchell, G., 2009. The relation between consistency and accuracy of eyewitness testimony: Legal versus cognitive explanations. In R. Bull, T. Valentine,

& T. Williamson (Eds.), *Handbook of psychology of investigative interviewing: Current developments and future directions* (pp.121–136). New York: Wiley-Blackwell.

Gioia, D. A., & Poole, P. P. (1984). Scripts in Organizational Behavior. *Academy of Management Review, 9*, 449-459. https://doi.org/10.5465/amr.1984.4279675

Granhag, P.A., Strömwall, L.A. & Hartwig, M., 2005. Granting asylum or not? Migration Board personnel's beliefs about deception. *Journal of Ethnic and Migration Studies*, *31*, 29–50. https://doi.org/10.1080/1369183042000305672

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 644-655. https://doi.org/10.1016/S0022-5371(74)80052-6

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth". *Perspectives on Psychological Science, 6*, 253-271. https://doi.org/10.1177/1745691611406924

Hope, L., Gabbert, F., Kinninger, M., Kontogianni, F., Bracey, A., & Hanger, A. (2019). Who said what and when? A timeline approach to eliciting information and intelligence about conversations, plots, and plans. *Law and Human Behavior*, *43*, 263-277. https://doi.org/10.1037/lhb0000329

Hope, L., Mullis, R., & Gabbert, F. (2013). Who? What? When? Using a timeline technique to facilitate recall of a complex event. *Journal of Applied Research in Memory and Cognition*, *2*, 20-24. https://doi.org/10.1016/j.jarmac.2013.01.002

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

Kaasa, S. O., Morris, E. K. & Loftus, E. F., 2011. Remembering why: Can people consistently recall reasons for their behaviour? *Applied Cognitive Psychology*, *25*, 35–42. https://doi.org/10.1002/acp.1639

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*, 103-109. https://doi.org/10.3758/BF03197276

Kelloway, E. K., Stinson, V., & MacLean, C. (2004). Eyewitness testimony in occupational accident investigations: Towards a research agenda. *Law and Human Behavior, 28*, 115-132. https://doi.org/10.1023/B:LAHU.0000015006.67141.44

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. Psychological Review, 103, 490-517. https://doi.org/10.1037/0033-295X.103.3.490

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*, 481-537.

Krinsley, K. E., Gallagher, J. G., Weathers, F. W., Kutter, C. J., & Kaloupek, D. G. (2003). Consistency of retrospective reporting about exposure to traumatic events. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, *16*, 399-409. https://doi.org/10.1023/A:1024474204233

Kuznetsova, A., Brockhoff, P. B., & Christensen; R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*, 1-26. https://doi.org/10.18637/jss.v082.i13

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*, 149-169. https://doi.org/10.1037/0278-7393.7.3.149

Lindsay, D. S. (2008). Source Monitoring. In H. L. Roediger, III (Ed.), *Cognitive Psychology of Memory: Vol. 2.* J. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 325-348). Oxford: 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). London: SAGE.

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*, 337-363. https://doi.org/10.1037/a0039036

MacLean, C. L., Brimacombe, C. A., & Lindsay, D. S. (2013). Investigating industrial investigation: Examining the impact of a priori knowledge and tunnel vision education. *Law and Human Behavior, 37*, 441-453. https://doi.org/10.1037/lhb0000056

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*, 474- 486. https://doi.org/10.1002/acp.3421

Minsky, M. (1974). *A framework for representing knowledge*. Cambridge, MA, USA: Massachusetts Institute of Technology.

Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience and Biobehavioral Reviews, 36*, 1640-1645. https://doi.org/10.1016/j.neubiorev.2012.03.001

Neisser, U. (1988). What is ordinary memory the memory of? In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 356-373). New York, NY: Cambridge University Press.

Odinot, G., Wolters, G., & van Giezen, A. (2013). Accuracy, confidence and consistency in repeated recall of events. *Psychology, Crime & Law*, *19*, 629-642. https://doi.org/10.1080/1068316X.2012.660152

Postman, L. (1971). Organisation and interference. *Psychological Review, 78*, 290-302. https://doi.org/10.1037/h0031031

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: The Official Journal of the Society for Applied Research in Memory and Cognition, 11*, 339-360. https://doi.org/10.1002/(SICI)1099-0720(199708)11:4<339::AID-ACP460>3.0.CO;2-O

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. (2016). Children who experienced a repeated event only appear less accurate in a second interview than those who experienced a unique event. *Law and Human Behavior, 40*, 362. https://doi.org/10.1037/lhb0000194

R Core Team. (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from 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*, 550-558. https://doi.org/10.1016/j.tics.2012.09.003

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*, 407-417. https://doi.org/10.1002/acp.3118

Roediger, H. L. III & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.xRobinson, 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). Dordrecht, NL: Springer.

Rubínová, E. (2020a). Chapter 4. Adults' memory for instances of self-experienced interactive repeated event. In *Schema and deviation effects in adults' memory for repeated events* [Doctoral dissertation] (pp. 82-102). https://researchportal.port.ac.uk/portal/en/theses/schema-and-deviation-effects-in-adults-memory-for-repeated-events(cb6e11d0-28f5-45a5-97c8-641225fcafe8).html

Rubínová, E. (2020b). Appendix B. Disabling rehearsal in a repeated event paradigm disabled schema establishment and nullified deviation effects. In *Schema and deviation effects in adults' memory for repeated events* [Doctoral dissertation] (pp. 118-126). https://researchportal.port.ac.uk/portal/en/theses/schema-and-deviation-effects-in-adults-memory-for-repeated-events(cb6e11d0-28f5-45a5-97c8-641225fcafe8).html

Rubínová, E., Blank, H., Koppel, J., & Ost, J. (2021). Schema and deviation effects in remembering repeated unfamiliar stories. *British Journal of Psychology, 112,* 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*, *28*, 309-322. https://doi.org/10.1080/09658211.2020.1712421

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart and the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the microstructures of cognition* (pp. 3-57). Cambridge, MA: MIT.

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*. https://doi.org/10.1016/j.jml.2019.104038

Schank, R. C. (1999). *Dynamic memory revisited*. Cambridge, UK: Cambridge University Press.

Sheridan, L. P., Blaauw, E., & Davies, G. M. (2003). Stalking: Knowns and unknowns. *Trauma, Violence, & Abuse, 4*, 148-162. https://doi.org/10.1177/1524838002250766

Slackman, E., & Nelson, K. (1984). Acquisition of an unfamiliar script in story form by young children. *Child Development*, 329-340. https://doi.org/10.2307/1129946

Stark, E., 2012. Looking beyond domestic violence: Policing coercive control. *Journal of Police Crisis Negotiations*, *12*, 199–217. https://doi.org/10.1080/15332586.2012.725016

Theunissen, T. P., Meyer, T., Memon, A., & Weinsheimer, C. C. (2017). Adult eyewitness memory for single versus repeated traumatic events. *Applied cognitive psychology, 31*, 164-174. https://doi.org/10.1002/acp.3314

Tryon, W.W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods, 6*, 371-386, https://doi.org/10.1037//1082-989X.6.4.371

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

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0. http://www.stats.ox.ac.uk/pub/MASS4/

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software*, *36*, 1-48. https://doi.org/10.18637/jss.v036.i03

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

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*, 206-225. https://doi.org/10.1177/0734016817704700

Woiwod, D. M., Fitzgerald, R. J., Sheahan, C. L., Price, H. L., & Connolly, D. A. (2019). A meta-analysis of differences in children's reports of single and repeated events. *Law and Human Behavior, 43*, 99-116. https://doi.org/10.1037/lhb0000312

Appendix A

Table A

Activities and details in four sets of stimuli used across visits

Activity		
Game (Color): Roll three dice and tell a story	Products (Brand/Type): Evaluate packaging design	Device (Tasks): Inspect and evaluate device
Original (Red)	Shower gels (Radox, Lynx, Sanex)	Vertical mouse (draw a house, car, and a tree in Paint)
Voyages (Green)	Sweet treats (Lindt Dark Chilli Chocolate, Maryland Cookies, Terry's Chocolate Orange,)	White marker (draw an exclamation mark, a tick sign, and a dollar sign)
Actions (Blue)	Deodorants (Nivea, Old Spice, Sure)	Laser pointer (circle around a monkey, underline a horse, and cross out a chicken)
Fantasia (Purple)	Soft drinks (Coconut Water, Ginger Beer, Monster Energy)	Decorative scissors (cut out a card, a heart, and a star)