Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# The effect of timing and content deviations on adults' memories of a repeated event

**Stefanie J. Sharman**

stefanie.sharman@deakin.edu.au

Deakin University

**Meaghan C. Danby**

Deakin University

**Celine van Golde**

The University of Sydney

**Justine Derbyshire**

Deakin University

**Article**

**Keywords:**

**Posted Date:** November 19th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11451862/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Abstract

Witnesses' reports of repeated events are improved when one of the episodes deviates from the others. The current experiment examined whether the similarity between the deviation and the rest of the episodes, and the timing of the deviation, affects script development and recognition memory. Two hundred and forty-seven participants watched four standard episodes depicting intimate partner violence; some watched a deviation episode that had high-, moderate-, or low-similarity to the standard episodes. The timing of the deviation episode was manipulated so that it was the first, second, third, fourth, or fifth episode that participants saw. Compared to participants that did not experience a deviation, deviation participants showed disrupted script formation at the time of the deviation episode. Low-similarity deviations were more disruptive than high-similarity deviations. However, deviations had little impact on recognition accuracy for the last standard episode. While deviation participants were more accurate for details that changed across the standard episodes compared to non-deviation participants, they were not more accurate for details that were fixed across episodes or new details that occurred in only one episode. Our results suggest that participants rapidly developed scripts, but this development was disrupted when a deviation occurred, particularly when it was very different from the other episodes. These findings extend current understanding of script development and have practical implications for witnesses' memory of a repeated event, where distinctive incidents may help individuals differentiate between episodes without necessarily improving memory for all event details.


# Introduction

Adults may experience crimes repeatedly before reporting them to police, such as family violence, sexual abuse, stalking, and workplace bullying [1]. When they report these repeated events to authorities, complainants need to provide as much accurate detail about each individual episode as they can [2, 3, 4]. Providing detailed information assists investigators in establishing the offence and the perpetrator(s), allows corroboration with external sources of information, such as physical evidence, and can make witnesses appear more credible [e.g., 5, 6].

Given the importance of memory for repeated events, researchers have taken an interest in understanding the factors that impact adults' reports of individual episodes of events such as intimate partner violence videos, workplace bullying scenes, food tasting sessions, and car accidents [7, 8, 9, 10]. This work has reliably shown that, compared to participants with a single event experience, those with repeated experience more proficiently recall details that do not change across episodes (*fixed* details, such as the main characters of the event), but are poorer at reporting the details that change each episode (variable details, such as the colour of the main character's t-shirt changing in each episode from blue, to red, purple, and green [8, 11]). Specifically, variable details may be recalled at two levels: a broad item-level detail can be well recalled (e.g., "he always wears a t-shirt"), but the particular instantiation-level (one time the t-shirt was blue) is often misattributed to an incorrect episode (saying it was blue in episode two when it was actually blue in episode one [see 8; for a review, see also 11]). This misattribution of variable instantiations is termed an internal intrusion error. Confusing details that were

never experienced in any episode (e.g., saying the t-shirt was pink, when this colour never occurred) is rare for repeated event participants; this is termed an external intrusion error [11].

One explanation proposed for these frequently occurring internal intrusion errors is based on people's source monitoring decisions. According to the *Source Monitoring Framework*, people evaluate the qualitative characteristics of their memories (such as the amount of perceptual detail, contextual information, and emotional content they contain) and apply heuristics or deliberate reasoning to judge the memory's source [12, 13]. Errors occur when these characteristics overlap across sources, such as when episodes of a repeated event contain similar details, or when judgments are not made systematically [12]. Characteristics that make one episode stand out from the others are likely to improve source monitoring decisions [13].

One factor that has been shown to increase accuracy when reporting episodes from a repeated event series, is when there is a deviation episode that differs from the standard episodes [9, 14, 15; see also 16 for similar findings in children]. For example, in Experiment 1 in one study [9], adult participants took part in five tasting sessions over 1 hour. Each session contained fixed details (e.g., participants always sat at the table) and variable details (e.g., participants always ate a vegetable, but the specific vegetable differed each session: cauliflower, cucumber, celery, carrots and broccoli). Details were always presented in the same temporal order. For half the participants, the third episode was a deviation episode; it was completed in a room that was different from all the other episodes. Participants that experienced the location deviation demonstrated improved recall of the variable details from the episode in which the deviation occurred, as well as their recall of variable details from all episodes in the repeated event. Participants may have spent time comparing similarities and differences between the deviation episode and the other episodes in the sequence, which led to deeper processing of variable details (compared to participants who did not receive a deviation event; [9]).

Improved recall for both the deviation and the standard episodes has been found in several deviation studies, using a variety of methods to differ the deviation episode from the others [14, 15, 16]. To develop an episode that deviates from the rest, studies have added a new detail, replaced typical details with something new, changed the order of typically experienced details, held a standard episode in a new location, or some combination of these factors [9, 14, 15, 16, 17]. Some evidence suggests that the nature of the deviation may impact memory. For example, in Rubínová et al.'s study, participants were exposed to lists of words that were presented three times to establish participants' expectations for content and order; that is, they saw animal words, followed by clothing words, followed by fruit words [14]. In the fourth and final list, participants experienced a content deviation (a new category of kitchen words), an order deviation (the sequence of categories was changed), both deviations, or no deviation (control condition). In a subsequent recall test, participants recalled as many words as they could from all four lists. Content deviations increased recall for the final list as well as for all lists (Experiment 2); order deviations had the opposite effect and decreased recall for the final list as well as for all lists (Experiment 1). In another study with child participants, 9- to 11-year-olds (but not 6- to 8-year-olds) had better recall of four episodes of a repeated magic show if they had experienced a deviation causing a

continual disruption to the entire episode, where every activity occurred differently, rather than a deviation to only one activity within an episode [16].

It is possible that deviation episodes with a significant difference (that is, they are low in similarity to the standard episodes) may lead to larger memory benefits than deviation episodes that are higher in similarity to the standard episodes [e.g., 16]. This may occur because low-similarity deviation episodes stand out more against the routine of the standard episodes. When people experience repeated episodes of an event, they develop a "script" or general memory representation of the event that includes information about what usually happens (such as fixed details) and the temporal order of events [18, 19]. Variable details are represented in scripts at an item-level (e.g., "I always eat a vegetable") as a slot in the script, waiting to be filled by each instantiation. Scripts direct adults' expectations for what to anticipate in future episodes but also provide a retrieval framework to recall past episodes [see 17, 20, 21]. Establishing a script frees up cognitive resources for people to better recognise atypical details that deviate from the script [22]. It follows that deviation episodes containing significant or plentiful differences from the standard episodes may be more recognisable in memory than episodes only mildly deviating from the standard episodes. This idea is supported by Rubínová and colleagues' findings; when asked, their participants were able to describe their expectations around the content and order of the word lists (i.e., their scripts) and most of the participants (75%) who experienced a content deviation noticed it and were able to describe the change, but only 40% of participants noticed the more minor deviation to the event order [14].

While they may improve memory for events with an established script, low similarity deviations may also disrupt people's development of a script for new repeated events. According to the Schema-Confirmation Deployment Model [17, 22], when people experience the first episode of a new repeated event, all the details are new and can be considered prospective script components. With exposure to each additional episode, people can confirm which details are typically occurring and integrate them into a script. This process is known as "schema-confirmation" and lasts until people have built up enough experience of the event to develop a script of what usually happens [22]. After confirmation, a script is then implemented ("deployed"), and atypical details (that is, details not consistent with the script—such as variable instantiations and content deviations) become easier to notice [22, 23]. While still in the script development stage, experiencing a slight deviation to the typical episodes may not be overly disruptive to script development, but experiencing a very different deviation episode may be more detrimental.

Once the script has been deployed, atypical details are not absorbed into the script but stored separately in a representation that is linked to the script [18; see also 24]. The additional processing these variable and deviation details require typically leads to better recall [e.g., 25, 26, 27]. Deviations may also encourage comparisons between episodes, which leads to deeper processing through establishing connections, and helps people to consolidate their memories for the variable details across episodes [9, 16]. This deeper processing should lead to better recall overall.

Given that script development occurs over several episode exposures, the timing of a deviation episode within the sequence of a repeated event may impact not only the development of the script but also a person's memory accuracy for an individual episode of the event. If participants have not experienced enough episodes of an event to confirm which details are typically occurring, they are unlikely to be able to pay more attention to atypical details (since all the details are new). In this situation, such as a deviation occurring immediately after the first episode of the repeated event, the deviation episode should not benefit memory accuracy. However, if the deviation episode occurs after several episodes of a repeated event, people may have developed and confirmed an event script and moved to the schema-deployment phase where they can devote more attention to processing atypical details, which should result in more accurate memory reports.

To date, the timing of a deviation episode has only been explored in children [22]. In this study, 4- and 7-year-old children took part in four episodes over a 1-week period in which they completed several activities. For three of the episodes, all details were fixed and occurred identically. In the remaining episode, children experienced a deviation to the content (a new activity and changes to an existing activity) and to the order (two activities were swapped). Half the children experienced the deviation episode as the first episode, and half experienced it as the last episode. One week after the last episode, children were interviewed about the standard and the deviation episodes. For 4-year-olds, recall of the script-inconsistent details from the deviation (such as the new activity) was significantly higher when the deviation occurred as the last, rather than first, episode. Deviation timing did not affect the 7-year-olds' report of script-inconsistent information. This pattern suggests that the older children could deploy a script (supporting their recall of script-inconsistent details) within the span of three standard episodes regardless of the deviation timing, but the younger children could not deploy their scripts when the deviation came prior to any standard episodes. Having the deviation first likely made it more difficult for them to confirm which details were typically occurring.

Adults can encode more detailed and complex scripts after fewer event repetitions than children [19]. Thus, they may be less impacted by the presence of an early deviation episode than young children. Indeed, in one study, participants were exposed to four versions of an unfamiliar story; they recalled each story immediately after its presentation, as well as at longer retention intervals [15]. Some participants experienced a deviation in the fourth story—either a content or order deviation or both. Participants' immediate recall of details showed the largest increase from Story 1 to Story 2, with only modest increases for later stories. This finding suggests that participants' scripts were largely developed through exposure to the first two stories and that they completed the schema-confirmation stage after being exposed to all four stories. However, participants had difficulty remembering the content deviations, and the order deviation reduced recall but only for the last episode. Further, participants' awareness of the deviations was relatively low. These findings suggest that participants had not yet reached the schema deployment phase and, as such, were unable to pay more attention to (and demonstrate better memory for) atypical details. The authors suggested that participants might not have reached schema deployment due to the unfamiliar nature of the stories which meant that it took longer for them to establish the script [15].

In the current experiment, we varied the timing of participants' exposure to a deviation episode to examine its impact on their script development as well as their memory. Participants experienced four standard episodes of a repeated event. Some participants also experienced a deviation episode. The timing was manipulated so that it was experienced before the first standard episode, before the second, third, or fourth. We manipulated the similarity of the deviation episode, so that it had high, moderate or low similarity with the standard episodes. Finally, within each standard episode, we manipulated the types of details that participants experienced. In addition to fixed and variable details, we introduced new details that were unique to each standard episode. Each standard episode contained two new details (one central and one peripheral), and these details did not appear in any other episode. For example, in the fourth episode, the female character feeds her leftovers to the cat (new central detail) and there is a yellow hatstand in the hallway when the male character comes home (new peripheral detail). These new details were included to explore if they helped participants differentiate between the different episodes. As they would not have been part of the script that participants developed, we were interested in whether these atypical details received additional processing and were better remembered.

We made several predictions about the development of participants' scripts and their recognition memory of the target episode (that is, the last standard episode). First, we predicted that the presence and timing of the deviation episode would affect participants' script development. Participants who did not experience a deviation episode were expected to develop their script over the four standard episodes; they should show a linear increase in the number of details they predicted to occur in each standard episode, with the biggest increases occurring early in the sequence [15]. For participants who were exposed to a deviation episode, they should show a similar linear increase, apart from when they experienced the deviation episode. The deviation episode should decrease participants' predicted details as their script is temporarily disrupted by the deviation. However, in the next standard episode, their predicted details should increase again.

Second, we predicted that the similarity of the content of the deviation episode to the standard episodes should affect participants' predictions. We expected that participants exposed to the high-similarity episode should not experience as much of a disruption to their script development as participants exposed to the low-similarity episode. Therefore, after the first two episodes, participants exposed to the high-similarity deviation should predict more details than participants exposed to the low-similarity deviation.

Our third prediction was for participants' recognition memory: we predicted that participants who experienced a deviation episode should show greater accuracy for the target episode (the last episode) than participants who did not. Fourth, it was predicted that the timing of deviation episode should affect participants' recognition accuracy; specifically, participants who experienced at least two standard episodes before the deviation episode should have developed at least a partial event script and some may have moved into the schema-deployment phase, allowing them to devote more attention to script atypical details. Thus, these "later deviation" participants may show better recognition accuracy for the new details than "early deviation" participants who had not experienced at least two standard episodes

before the deviation episode. We did not make any predictions about participants' confidence ratings for their recognition test responses.

## Method

# Design and Participants

The experiment had a 3 (Deviation content: low-similarity, moderate-similarity, high-similarity) x 5 (Deviation timing: first, second, third, fourth, fifth) between-subjects design. An *a priori* power analysis, with power = .80 and α = .05, showed that a sample size of 103–249 participants was needed to detect an interaction depending on whether the effect size was medium ($f$ = 0.25) or large ($f$ = 0.40). Two-hundred forty-seven adults aged 18–65 years ($M$ = 36.56, $SD$ = 14.50) completed the study. One-hundred sixty-one identified as female, 84 identified as male, and two identified as non-binary.

## Materials

# Standard episodes

Participants watched four videos that followed the same story arc depicting intimate partner violence; each video was between 4 min 15 sec and 4 min 57 sec long. The female character (Sarah) is in the kitchen when her husband (John) comes home from work. After he changes into casual clothes, they sit down for dinner. During the conversation while they are eating, John takes offense at something Sarah says. He becomes aggressive and verbally and/or physically abuses her. Sarah goes to the bathroom to recover. When she returns, they go and sit on the couch and watch TV. The episodes contained seven fixed details that did not change across videos: six were contextual details including the two central characters (John and Sarah), the location (their house), Sarah's clothes, Sarah's hairstyle, and John's clothes. The final fixed detail was part of the story arc and showed the ending to the episode: Sarah and John always sat on the couch and watched TV. Each video also contained 10 variable details that had a different presentation in each episode, and two new details that never recurred in any other episode (see Table 1). All variable details were central to the story arc. Of the two new details in each episode, one was a central detail that the characters interacted with during the video, the other was a peripheral detail that appeared in the background of one of the scenes in the video. The main difference across episodes is that the level of violence escalates from yelling in episode 1 through to physical abuse in episode 3 (grabbing Sarah's wrist) and episode 4 (grabbing the back of her neck and head); this escalation mimics real-life abuse [e.g., 28].

## Deviation episodes

Three versions of a deviation episode were developed that differed in similarity to the standard episodes. All versions contained the same two characters (Sarah and John) and were designed to be negative in

tone. In the high-similarity episode (4 min 55 sec), Sarah and John are in the same location (their house) and John verbally and physically abuses Sarah. The main difference to the standard episodes is that the high-similarity episode is set in the morning and follows John and Sarah through their morning routine, rather than being set at night around dinner. In the moderate-similarity episode (4 min 36 sec) and the low-similarity episode (4 min 38 sec), Sarah and John are in a different location (a park), and the videos are set around a lunchtime picnic rather than dinner time. In the moderate-similarity episode, John verbally abuses Sarah during a fight about her speaking to another man. In the low-similarity episode, there is no abusive event but rather Sarah remembers her grandmother who passed away,

Table 1

*Variable and new items to which participants were exposed in the standard episodes*.

| | **Episode 1** | **Episode 2** | **Episode 3** | **Episode 4** |
|---|---|---|---|---|
| **Variable items** | | | | |
| Sarah in kitchen at start | Washing sink | Pot on stove | Looks in oven | Loading dishwasher |
| Music playing | Pop | Jazz | Rap | Heavy metal |
| John arrives home carrying an item | Lunch container | Brought milk | Mail | Newspaper |
| John's t-shirt colour | Blue | Orange/peach | Purple | White |
| Drink with dinner | Wine | Beer | Coke | OJ |
| Dinner | Sausages and mash | Spaghetti & meatballs | Pizza | Fish and chips |
| Trigger/fight topic | Spending money/buying shoes for wedding | Jealousy of co-worker (Simon) | Too much time with friends | She badmouths him about not going for a promotion |
| Abusive/aggressive act | Looks furious/ throw napkin / calls her gold-digger | Yelling | Grab wrist | Grab back of head/neck/hair |
| Bathroom activity | Texting | Crying | Wash hands and dab water on face | Fix hair |
| Him doing something when she's in bathroom | Finish off her drink | Looks through her bag | Taking plates to kitchen | Look at his own phone |
| **New items (central)** | | | | |
| Food needs seasoning | Salt added | | | |
| Table centrepiece | | Silver candle stick holder | | |
| Dessert | | | Chocolate cake | |
| Woman feeds leftovers to pet | | | | Cat ("Bella") |
| **New items (peripheral)** | | | | |

| | Episode 1 | Episode 2 | Episode 3 | Episode 4 |
|---|---|---|---|---|
| **Variable items** | | | | |
| Chairs | High back wooden dining chairs | | | |
| Hand towel | | Orange towel (hanging on hook) | | |
| Item left on couch (in background of dinner) | | | Laundry basket | |
| Hallway item | | | | Yellow hatstand |

so that the overall emotional tone of the episode was slightly negative. The top half of Table 2 summarises the main consistencies and differences between the three deviation episodes and the standard episodes.

Table 2
*Consistencies and differences between the deviation and standard episodes.*

| **Sarah & John** | **High similarity** | **Moderate similarity** | **Low similarity** |
|---|---|---|---|
| | ***n* = 75** | ***n* = 78** | ***n* = 76** |
| | **Present** | **Present** | **Present** |
| Location | Same as standard episodes | Different to standard episodes | Different to standard episodes |
| Abuse | Occurs | Occurs | Does not occur |
| Tone ratings | 1.47 (0.55)[a] | 1.72 (0.74)[a] | 5.99 (1.09)[b] |
| Happiness | 1.65 (0.63)[a] | 1.81 (0.79)[a] | 5.21 (1.25)[b] |
| Comfort | 1.81 (0.77)[a] | 1.79 (0.84)[a] | 5.17 (1.35)[b] |
| Example of IPV | 6.65 (0.73)[a] | 6.19 (0.94)[c] | 1.78 (1.10)[b] |

Notes: IPV = Intimate Partner Violence; ratings were made on 7-point Likert scales, with high scores presenting more positive ratings. [a]Means with the same superscript letter are not significantly different ($p > .05$).

## Recognition tests

There were two recognition tests: one for the target episode that was completed by all participants and one for the deviation episode, which was completed by participants who were exposed to a deviation

episode. For both recognition tests, yes/no questions were presented in the same order for all participants. The response options of “yes” and “no” appeared in random order for each question. For all recognition questions, participants rated their confidence in their answer from 1 (*not at all confident*) to 6 (*extremely confident*).

**Target episode.** The recognition test for the last standard episode (the target episode) consisted of 24 questions. Four questions were about fixed details: three were contextual (Sarah’s clothes, John’s clothes, Sarah’s hairstyle), and one was about the episode ending (sitting on the couch and watching TV). Two questions had “yes” for the correct answer (e.g., “Did John arrive home wearing a suit in the [target episode]?”) and two questions had “no” for the correct answer (e.g., “Was Sarah wearing her hair down in the [target episode]?”).

There were 10 questions about variable details; one question for each variable detail. Two questions had “yes” as the correct answer as they presented the correct instantiation (e.g., “Was Sarah dancing to *heavy metal music* at the start of the [target episode]?”). The other eight questions had “no” as the correct answer. To measure internal intrusion errors for variable details, for six of these questions, the presented instantiation was from a different episode (e.g., “Did Sarah and John *drink wine* in the [target episode]?” which was incorrect as they drank orange juice in the target episode and wine in episode 1). The incorrect instantiation presented was equally from episode 1, episode 2, and episode 3 across these six questions. To measure external intrusion errors for variable details, for the remaining two questions the presented instantiation was not from any episodes (e.g., “Did John change into a *black* t-shirt in the [target episode]?” when this colour never occurred in any video). In sum, 10 questions assessed accuracy for variable details; within these questions, two assessed accurate instantiations, six assessed internal intrusion errors and two assessed external intrusion errors.

There were 10 questions about new details; five were about central details and five were about peripheral details. Of the five questions about new central details, one question had “yes” as the correct answer (e.g., “Did Sarah scrape her leftover food into a pet bowl in the [target episode]?”). The other four questions had “no” as the correct answer. To measure internal intrusions for new central details, for three of these questions the presented new central detail was from a different episode (e.g., “Was there a silver candle centrepiece on the table in the [target episode]?”). To measure external intrusions for new central details, for one question the presented new detail was not from any episode (e.g., “Did John spill his drink on the table in the [target episode]?”). The five questions about new peripheral details followed the same structure: one had “yes” and four had “no” as the correct answer. Of those that had “no” as the correct answer, three questions presented new peripheral details from a different episode to measure internal intrusion errors, and one presented a new detail that was not from any episode to measure external intrusion errors. In sum, 10 questions assessed accuracy for new details; within these questions, two assessed accurate instantiations, six questions assessed internal intrusion errors (three for new central details and three for new peripheral details), and two questions assessed external intrusion errors (one for new central and one for new peripheral details).

**Deviation episode.** The recognition test for the deviation episode had three versions; one for each deviation episode. All recognition tests consisted of 12 questions. For six of the questions, the correct answer was “yes” and for the other six the correct answer was “no”. There were four questions about fixed details that included the location of the event (“Was most of the [deviation episode] set in Sarah and John’s home?”), and Sarah and John’s clothes and hairstyles (e.g., “Was John wearing a red flannel shirt in the [deviation episode]?”). The other eight questions were about new details that occurred in the episode. For example, for the low- and moderate-consistency episodes, the questions were about the car in which John and Sarah arrived at the picnic area, the colour of their picnic rug, and the reason for Sarah being upset at the end of the episode. For the high consistency episode, the questions included the time that their alarm was set for in the morning, the colour of the bathroom mat, and the reason for Sarah being upset at the end of the episode. In sum, four questions assessed accuracy for fixed details, and six questions assessed accuracy for new details.

# Procedure

### Ethics approval

was granted from Deakin University (HREC 2023-039). The research was performed in accordance with relevant guidelines and regulations. Participants responded to advertisements placed around the university campus and to emails sent by the researchers. The study was advertised as examining adults’ emotional responses to intimate partner violence and no mention of memory was made. This cover story allowed us to inform participants of the nature of the videos that they would be watching without warning them about the memory component, as we did not want participants to pay more attention to the videos than they otherwise would, or to make notes or record any aspects of the videos. After providing informed consent, participants took part in two sessions.

**Session 1.** Session 1 was conducted wholly online in a location of participants’ choosing. Participants provided demographic information before watching the four standard episodes. They were randomly assigned to watch one of the deviation episodes or no deviation episode. If they watched a deviation episode, the timing of the episode was randomly assigned such that they watched it first (before the first standard episode), second (before the second standard episode), third (before the third standard episode), fourth (before the fourth standard episode), or fifth (after the fourth standard episode). After watching each standard or deviation episode, participants rated the tone of the video from 1 (*very negative*) to 7 (*very positive*), how happy they felt after watching the video from 1 (*very unhappy*) to 7 (*very happy*), how comfortable they felt after watching the video from 1 (*very uncomfortable*) to 7 (*very comfortable*), and whether they thought what they saw in the video was an example of intimate partner violence (IPV) from 1 (*absolutely not*) to 7 (*absolutely yes*).

For the standard episodes, participants’ average ratings showed that they viewed the episodes as very negative ($M$ = 1.51, $SD$ = 0.44), they felt unhappy ($M$ = 1.72, $SD$ = 0.61) and uncomfortable ($M$ = 1.74, $SD$ = 0.68) after watching them. They thought that the episodes were an example of IPV ($M$ = 6.20, $SD$ = 0.83).

Repeated measures ANOVAs with episode as the repeated factor showed that as they watched each standard episode, participants' ratings got increasingly negative, more unhappy and more uncomfortable, *F*s > 132.16, *p*s < .001, $\eta_p^2$ > .350, and they increasingly rated the episodes as examples of IPV, $F(1, 244) = 219.17$, $p < .001$, $\eta_p^2 = .473$.

For the deviation episodes, participants' ratings varied across the three episodes (see the lower half of Table 2). Due to violations of sphericity and homogeneity of variance, separate Welch ANOVAs were conducted. For all ratings, there was a significant effect for deviation episode, Welch *F*s > 199.97, *p*s < .001, $\eta_p$s> .709. Bonferroni-corrected posthoc tests revealed that the low-similarity episode had higher tone, happiness, and comfort ratings than both the moderate- and high-similarity episodes, all *p*s < .001. For IPV ratings, participants gave lower ratings to the low-similarity episode than both the moderate- and high-similarity episodes. They also gave lower ratings to the moderate- than the high-similarity episodes, $p = .008$.

To measure the development of participants' scripts for what usually happens in the episodes, we asked them to predict what the next video would show immediately after watching each one. For example, after watching the first video, they were told, "Please describe everything you think will happen in the next video, from beginning to end. Please be as specific as possible in your response." Participants were provided a free-text box for their response. After completing their response, participants completed a distractor task for 1 minute. Next, participants watched the second standard episode (or deviation episode). They completed the same ratings of tone, happiness, comfort, and whether the video showed an example of intimate partner violence, before predicting what they thought would happen in the next video. After completing their responses, participants completed a distractor task for 1 minute. This sequence repeated until participants had watched all four standard episodes, and one deviation episode (for those that were exposed to a deviation episode).

**Session 2.** Three days after Session 1, participants completed Session 2. Session 2 was conducted face-to-face at a location of the participant's choosing. Most participants had a three-day interval ($M = 2.97$, $SD = 0.24$); however, rescheduling meant that some participants had a two-day ($n = 11$) or four-day ($n = 3$) interval. Interval was not related to any of our dependent variables, *r*s < .06, *p*s > .364. Session 2 was conducted by one of 16 research assistants; the research assistant who conducted the session was not related to the participants' recognition accuracy, $F(15, 231) = 1.55$, $p = .089$, $\eta_p^2 = .091$. First, research assistants spent 1–2 minutes building rapport with the participant by asking questions such as, "How's your day been going so far?" They were instructed to share a little information about themselves to participants to establish a reciprocal relationship see [29]. Then, participants responded to free recall questions before completing a recognition questionnaire. As the free recall data showed similar patterns to the recognition data, they are not presented here or described further.

Participants completed an online recognition test using the research assistant's laptop. They were asked recognition questions about the target episode and the deviation episode (if they were exposed to one). For participants who answered recognition questions about both the target episode and the deviation

episode the order of questions was counterbalanced, such that 118 participants were asked about the target episode first and 111 were asked about the deviation episode first. Participants who were asked about the target episode first were cued with the statement, “First, I am going to ask you about the [last] video you watched, showing John and Sarah having dinner.” Participants who were asked about the deviation episode first were cued with the statement, “First, I am going to ask you about the video in which John and Sarah had a picnic/got up in the morning.” Whether they were asked about the target or deviation episode first had no effect on recognition accuracy for either the target episode, $F(1, 227) = 1.36$, $p = .245$, $\eta_p^2 = .006$, or the deviation episode, $F(1, 226) = 0.39$, $p = .532$, $\eta_p^2 = .002$.

## Coding and analysis

Participants’ predictions about what would happen in the next episode were coded from Session 1. Their recognition responses were coded from Session 2.

**Script predictions.** Participants’ sets of predictions of what would happen in the next episode were coded for 11 typically-occurring details. These 11 details focused on the main action sequences in the standard episodes; that is, the 10 variable item-level details and the ending of the episode (which was fixed and showed the two main characters sitting on the couch watching TV). Participants were awarded 1 point for each of the 11 details that they correctly predicted; these points were summed to represent the number of correct details that participants predicted would occur in the next episode (*range* = 0–11). One coder coded all participants’ script predictions; a second coder coded 20% for reliability. Inter-rater reliability was strong, Cohen’s kappa = .85.

**Recognition responses.** For the recognition test about the target episode, participants’ responses for questions about fixed details were coded as correct (1) or incorrect (0). Their proportion accuracy was calculated for fixed details by summing the number of correct responses and dividing this sum by the total number of questions (4).

For questions about variable details, participants’ responses were coded as correct (1) or incorrect (0). Their proportion accuracy was calculated for variable details by summing the number of correct responses and dividing this sum by the total number of questions (10). For the six questions that also measured internal intrusion errors, participants’ incorrect responses were summed and divided by six to calculate their proportion of internal intrusions. For the two questions that also measured external intrusion errors, participants’ incorrect responses were summed and divided by two to calculate their proportion of external intrusions.

For questions about new central and peripheral details, participants’ responses were coded as correct (1) or incorrect (0). Their proportion accuracy was calculated for new central details by summing the number of correct responses and dividing this sum by the total number of questions (5). Their proportion accuracy was calculated for new peripheral details by summing the number of correct responses and dividing this sum by the total number of questions (5). For the questions that also measured internal

intrusion errors, participants' incorrect responses were summed and divided by three to calculate their proportion of internal intrusions separately for new central details and new peripheral details. For the questions that also measured external intrusion errors, participants' incorrect responses were summed and divided by one to calculate their proportion of external intrusions separately for new central details and new peripheral details.

For the recognition test about the deviation episode, participants' responses to the questions were coded as correct (1) or incorrect (0). Their proportion accuracy was calculated for these details by summing the number of correct responses and dividing this sum by the number of questions (12).

## Results

Our dataset is available at: https://osf.io/h5u7j/overview?view_only=8f78b723470943878c97d183f7c5650b. The results are presented in two sections below. First, the results for participants' predictions about what they would see in the next episode are reported. Second, participants' recognition accuracy for target episode and the deviation episode are presented, along with their confidence ratings for their recognition test responses.

# Script predictions

We first examined participants' predictions about what they thought would happen in the next video. To determine whether the number of predicted details increased with each episode, we examined control participants who did not experience a deviation episode. Control participants predicted an average of 1.33 details (*SD* = 0.84) after watching the first episode, 2.78 details (*SD* = 2.63) after watching the second episode, 4.06 details (*SD* = 3.13) after watching the third episode, and 4.22 details (*SD* = 3.37) after watching the fourth episode. A repeated measures ANOVA[1] with episode as the repeated factor was significant, $F(2, 34) = 10.41$, $p < .001$, $\eta_p^2 = .380$. Within-subjects contrasts revealed the linear contrast was significant, indicating that predicted details increased linearly with greater exposure to the episodes, $F(1, 17) = 15.62$, $p = .001$, $\eta_p^2 = .479$.

We then examined participants who experienced a deviation episode to determine its effect on participants' predicted details and whether the content and timing of the deviation episode mattered. Figure 1 shows the number of details participants predicted would occur in the next episode after watching each one. The figure shows that, in general, across the five episodes (four standard and one deviation), participants' predicted details increased with an increasing number of episodes viewed; however, the number decreased for predictions made immediately after viewing the deviation video. After another standard episode, participants' predicted details began to increase again.

Due to assumptions of homogeneity of covariance and variance being violated, separate one-way Kruskal Wallis non-parametric tests were used to determine the effects of timing of the deviation episodes (deviation as episode 1, 2, 3, 4 or 5), and of the deviation content (low, moderate, or high

similarity) on participants' predicted details. For predictions made after viewing the video in episode 1, a Kruskal Wallis test showed no significant effect of deviation timing or content, $H(4) = 9.82$, $p = .057$, $\eta^2 = .026$, and $H(2) = 0.31$, $p = .859$, $\eta^2 = .008$, respectively.

For predictions made after watching the video in episode 2, there was a significant effect of deviation timing, $H(4) = 50.32$, $p < .001$, $\eta^2 = .208$. Bonferroni-corrected pairwise comparisons revealed that participants predicted significantly more details if they were to experience the deviation in episodes 3, 4, and 5 (i.e., the deviation was not yet experienced) than if it was experienced in episode 1, *p*s < .005, or in episode 2, *p*s < .005. No other differences were significant, *p*s = 1.00. There was no significant effect of deviation content on predicted details, $H(2) = 1.37$, $p = .504$, $\eta^2 = .003$.

For predictions made after watching the video in episode 3, there was a significant effect of deviation timing, $H(4) = 40.79$, $p < .001$, $\eta^2 = .165$. Bonferroni-corrected pairwise comparisons revealed that participants predicted significantly more details if the deviation was to occur in episodes 4 and 5 (i.e., the deviation was not yet experienced) than if it had been experienced as episode 3, *p*s < .001, or as episode 2, *p*s < .019. They also predicted more details if the deviation was to occur as episode 5 rather than if it had been experienced in episode 1, $p = .002$. No other differences were significant. There was a significant effect of deviation content, $H(2) = 11.13$, $p = .004$, $\eta^2 = .041$. Participants who watched the high-similarity episode predicted more details than those who watched the low-similarity episode, $p = .003$. No other differences were significant, *p*s > .232.

For predictions made after watching the video in episode 4, there was a significant effect of deviation timing, $H(4) = 41.51$, $p < .001$, $\eta^2 = .168$. Bonferroni-corrected pairwise comparisons revealed that participants who had just seen the deviation as episode 4, predicted significantly fewer details than if the deviation had occurred (or was yet to occur) in any other episode, *p*s < .014. Further, participants who were to see the deviation in episode 5 (i.e., they had not yet experienced it) predicted significantly more details than participants who had experienced the deviation episode as episode 1, 2, or 3, *p*s < .018. No other differences were significant. There was an effect of deviation content, $H(2) = 11.56$, $p = .003$, $\eta^2 = .042$. Participants who watched the high-similarity episode predicted more script details than those who watched the low-similarity episode, $p = .003$. No other differences were significant, *p*s > .051.

For predictions made after watching the video in episode 5, there was a significant effect of deviation timing, $H(4) = 15.53$, $p = .004$, $\eta^2 = .52$. Bonferroni-corrected pairwise comparisons revealed participants predicted significantly fewer details if they had just viewed the deviation episode as episode 5 than those who viewed the deviation as episodes 1, 2, or 3, *p*s < .037. No other differences were significant, *p*s > .419. There was an effect of deviation content, $H(2) = 6.89$, $p = .032$, $\eta^2 = .022$. Participants who watched the high similarity episode predicted more script details than those who watched the low similarity episode, $p = .032$. No other differences were significant, *p*s > .207.

# Recognition accuracy and confidence

# Deviation episode

**Accuracy.** Before examining participants' recognition accuracy for the target episode, we first examined their recognition accuracy for the deviation episode to ensure that they paid attention to the episode and to explore whether there were any differences between the three deviation episodes. To ensure that they paid attention to the deviation episode, participants' recognition accuracy was compared to chance (.50 as there were two options for each of the recognition questions). Overall, participants' recognition accuracy for the deviation episode ($M$ = .69, $SD$ = .13) was significantly higher than chance, $t(227)$ = 22.06, $p$ < .001, Cohen's $d$ = 1.46.

Next, to determine if their accuracy was affected by the similarity of the deviation episode to the standard episodes, a one-way ANOVA (similarity: high, moderate, low) was conducted. There was a significant main effect for similarity on recognition accuracy, $F(2, 225) = 3.64$, $p$ = .028, $\eta_p^2$ = .031. Bonferroni-corrected posthoc tests revealed that participants were more accurate for the moderate-similarity episode ($M$ = .72, $SD$ = .13) than the high-similarity episode ($M$ = .67, $SD$ = .14), $p$ = .028. There was no difference in recognition accuracy between the low-similarity episode ($M$ = .69, $SD$ = .13) and the other deviation episodes, $p$s > .200.

**Confidence.** We examined participants' confidence in their recognition responses for the deviation episode. A one-way ANOVA (similarity: high, moderate, low) could not be conducted due to the assumption of homogeneity of variances not being met. A Kruskall-Wallis nonparametric test revealed a significant main effect for similarity on confidence, $H(2) = 6.17$, $p$ = .046, $\eta^2$ = .019. Bonferroni-corrected pairwise comparisons revealed that participants were more confident for the moderate-similarity episode ($M$ = 3.93, $SD$ = 0.90) than the high-similarity episode ($M$ = 3.57, $SD$ = 0.84), $p$ = .040. There was no difference in confidence between the moderate- and high-similarity episodes and the low-similarity episode ($M$ = 3.70, $SD$ = 1.02), $p$s > .503.

# Target episode

**Deviation vs. no deviation.** We next examined participants' recognition accuracy for the target episode. First, we determined whether participants who were exposed to a deviation episode were more accurate than those who were not exposed. Four separate one-way ANOVAs were conducted to determine whether the presence of a deviation episode affected accuracy for each detail type (fixed, variable, new central and new peripheral). The top four rows of Table 3 shows participants' accuracy for each detail type, along with the corresponding ANOVA results. The presence of a deviation episode only affected proportion accuracy for variable details: participants were more accurate when a deviation was present than when it was not.

Overall, the presence of a deviation episode did not affect their confidence in their recognition test responses, Welch's $F(1, 25) = 0.11$, $p$ = .744, $\eta_p$ < .001. Participants were equally as confident when they experienced a deviation event ($M$ = 4.12, $SD$ = 0.88) as when they did not ($M$ = 4.07, $SD$ = 0.55). A Greenhouse-Geisser corrected repeated measures ANOVA was conducted to determine if participants'

confidence differed across the four detail types; the main effect for detail was significant, $F(3, 646) = 246.47$, $p < .001$, $\eta_p^2 = .500$. Bonferroni-corrected pairwise t-tests ($\alpha = .008$) showed that participants were less confident about new peripheral details ($M = 3.15$, $SD = 1.16$) than fixed ($M = 4.19$, $SD = 1.16$), new central ($M = 4.36$, $SD = 0.96$), and variable details ($M = 4.45$, $SD = 0.90$), all $p$s < .001. They were also less confident about fixed details than new central and variable details, all $p$s < .003. Participants were equally confident about new central and variable details, $p > .008$.

Next, we explored participants' internal and external intrusion errors. The middle section of Table 3 shows that for internal intrusion errors, there was no significant difference between participants exposed to a deviation and those who were not exposed, for any detail type. The lower section shows that for external intrusion errors, participants who experienced a deviation episode made fewer errors for variable details than those who did not experience one. There were no differences for new central or new peripheral details.

**Timing and similarity of deviation.** Our next analyses focused only on participants that experienced a deviation episode. We examined the effects of the timing of the deviation episode on participants' recognition accuracy by collapsing them into two groups: participants who experienced an earlier deviation (in the first two episodes) and those who experienced a later deviation (in the last three episodes). We also examined the effect of deviation similarity. Separate 2 (timing: early, late) x 3 (similarity: low, moderate, high) ANOVAs were conducted for each detail type.

For fixed details, there was no significant main effect for timing, $F(1, 223) = 2.42$, $p = .121$, $\eta_p^2 = .011$; participants' accuracy was similar regardless of whether they experienced a deviation early ($M = .76$, $SD = .22$) or late ($M = .71$, $SD = .23$). There was also no main effect for similarity (low-similarity $M = .74$, $SD = .22$; moderate-similarity $M = .72$, $SD = .25$; high-similarity $M = .74$, $SD = .20$), $F(2, 223) = 0.43$, $p = .654$, $\eta_p^2 = .004$. The interaction was also not significant, $F(2, 223) = 0.64$, $p = .528$, $\eta_p^2 = .006$.

Table 3
Accuracy, internal and external intrusion errors for fixed, variable and new details

| **Detail type** | **No deviation** *n* **= 18** | **Deviation** *n* **= 229** | ***F*(1, 245)** | ***p*** | $\eta_p^2$ |
|---|---|---|---|---|---|
| Proportion accuracy | | | | | |
| Fixed | .75 (.21) | .73 (.22) | 0.13 | .720 | .001 |
| Variable | .59 (.15) | .68 (.17) | 4.88 | .028 | .020 |
| New (central) | .71 (.21) | .66 (.21) | 1.09 | .297 | .004 |
| New (peripheral) | .50 (.17) | .58 (.22) | 2.27 | .134 | .009 |
| Proportion internal intrusion errors | | | | | |
| Variable | .46 (.23) | .39 (.23) | 1.91 | .169 | .008 |
| New (central) | .33 (.28) | .45 (.30) | 2.57 | .110 | .010 |
| New (peripheral) | .63 (.25) | .50 (.29) | 3.34 | .069 | .013[a] |
| Proportion external intrusion errors | | | | | |
| Variable | .42 (.35) | .26 (.31) | 4.32 | .039 | .017 |
| New (central) | .06 (.24) | .11 (.31) | 0.51 | .477 | .002 |
| New (peripheral) | .00 (.00) | .08 (.27) | 1.52 | .218 | .006[a] |

Note: Figures in parentheses are standard deviations. [a]Non-parametric tests were conducted due to the assumption of homogeneity of variance being violated; these tests showed the same pattern of results. The parametric tests are presented here for ease of interpretation.

For variable details, there was no significant main effect for timing, $F(1, 223) = 0.59$, $p = .121$, $\eta_p^2 = .011$; participants' accuracy was the same when they experienced a deviation early ($M = .67$, $SD = .19$) and late ($M = .69$, $SD = .17$). There was also no main effect for similarity, (low-similarity $M = .69$, $SD = .17$; moderate-similarity $M = .68$, $SD = .18$; high-similarity $M = .68$, $SD = .17$), $F(2, 223) = 0.08$, $p = .921$, $\eta_p^2 = .001$. There was no significant interaction, $F(2, 223) = 0.70$, $p = .489$, $\eta_p^2 = .006$.

For new central details, there was a trend towards an effect for timing, $F(1, 223) = 3.61$, $p = .059$, $\eta_p^2 = .016$. Participants were slightly less accurate when they experienced a deviation early ($M = .63$, $SD = .21$) than when they experienced it late ($M = .68$, $SD = .21$). There was no main effect for similarity (low-similarity $M = .67$, $SD = .21$; moderate-similarity $M = .66$, $SD = .20$; high-similarity $M = .65$, $SD = .22$),, $F(2, 223) = 0.24$, $p = .785$, $\eta_p^2 = .002$. The interaction was not significant, $F(2, 223) = 0.23$, $p = .794$, $\eta_p^2 = .002$.

Finally, for new peripheral details, there was no significant main effect for timing, $F(1, 223) = 1.56$, $p = .214$, $\eta_p^2 = .007$; participants' accuracy was the same when they experienced a deviation early ($M = .56$, $SD = .22$) and late ($M = .59$, $SD = .21$). There was also no main effect for similarity (low-similarity $M = .56$, $SD = .23$; moderate-similarity $M = .60$, $SD = .22$; high-similarity $M = .58$, $SD = .20$), $F(2, 223) = 0.75$, $p = .475$, $\eta_p^2 = .007$, and no interaction, $F(2, 223) = 0.69$, $p = .505$, $\eta_p^2 = .006$.

To determine if the similarity and timing of the deviation episode affected participants' internal and external intrusion errors, we conducted separate 2 (timing) x 3 (similarity) ANOVAs for variable, new central and new peripheral details. For internal intrusion errors, there were no significant main effects or interactions for any detail type, $F$s < 2.96, $p$s > .087. For external intrusion errors, there were also no significant main effects or interactions for any detail type, $F$s < 2.36, $p$s > .097.

**Confidence**. Finally, for participants that received a deviation episode, we explored their confidence for their responses on the recognition test. Confidence did not vary by deviation similarity, Welch's $F(2, 150) = 0.23$, $p = .792$, $\eta_p = .002$, or by deviation timing, $F(1, 227) = 1.11$, $p = .293$, $\eta_p^2 = .005$. However, it did vary by detail type. The pattern of results replicated the pattern above from the total sample and will not be repeated here.

## Discussion

In the current experiment, we examined the effect of timing and similarity of a deviation episode on participants' script development and memory for a repeated event. Consistent with our first hypothesis, participants who were not exposed to a deviation episode showed a linear increase in predicted details across the standard episodes, suggesting that their script for the repeated event built over time. They showed the greatest increases in predicted details from episode 1 to episode 2, and from episode 2 to episode 3. There was only a modest increase in predicted details from episode 3 to episode 4. These findings are consistent with past research showing that participants' immediate recall of details from repeated stories showed the largest increase from Story 1 to Story 2, with only modest increases for later stories [15]. Together, these findings suggest that adult participants need to experience two to three standard episodes of a repeated event to build a script while in the schema-confirmation phase [22].

When participants were exposed to a deviation episode, both the timing of the deviation episode and its similarity with the standard episodes affected their predicted script details. When participants experienced a deviation episode, they immediately predicted fewer script details for the next episode, likely because their script formation was disrupted by experiencing an episode that was different to the standard ones. Consistent with our second hypothesis, the level of disruption depended on the similarity between the deviation episode and the standard episodes: participants who experienced a high-similarity episode predicted more correct details than those who experienced a low-similarity episode. It is likely that the high-similarity episode, which contained many of the same features as the standard episodes, did not disrupt participants' script formation as much as the low-similarity episode, which shared fewer features with the standard episodes. This finding is consistent with the results of a study with children

[30]. In this study, 5-9-year-old children participated in four episodes across 2 weeks that had a high or low level of similarity. High-similarity episodes contained more fixed details, whereas low-similarity episodes contained more variable and new details. Children were interviewed about the episode they remembered best 5–8 days later. Results showed that high-similarity children reported more fixed details from their scripts and more new details atypical to the script than low-similarity children, suggesting that their scripts were confirmed and deployed faster than low-similarity children.

This finding extends the current literature to demonstrate the importance of timing and content of a deviation episode in script formation. Previous research has focused on changing the content or order of an episode, but this has comprised a smaller change within an otherwise standard episode. For example, in one study, the location of one of the taste-testings was changed in the deviation episode; but the taste-tastings experienced during the episode remained the same as the standard episodes [9]. In the story-based experiment, the content deviation meant that two of the 15 details were changed, but 13 remained unchanged [15]. In the order deviation, the content was the same, but the presentation order for 11 of the 15 details was changed. Our research shows that when the deviation episodes are even less similar to the standard episodes, such as in our low-similarity condition, these episodes are even more disruptive to participants' script formation.

The effect of a deviation episode (vs no deviation) on participants' recognition memory for the target episode showed mixed results. While the presence of a deviation improved accuracy for variable details and reduced external intrusion errors for these details, it did not affect accuracy for fixed details, new central details or new peripheral details. Thus, our third hypothesis was only partially supported. The finding for variable detail accuracy is consistent with the taste-testing experiment [9] and Connolly et al.'s experiment with children [16]. In the current study, the deviation episode appeared to help participants remember that a particular instantiation occurred (as shown by their improved accuracy overall) rather than did not occur (as shown by their reduced external intrusion errors). However, the deviation did not reduce internal intrusion errors, which suggests that it did not help participants associate each instantiation with the particular episode in which it occurred. It is possible that our participants who were exposed to the deviation episode spent more time thinking about and comparing the standard episodes with the deviation episode [see 9], which helped them to more accurately distinguish between the instantiations that were experienced and those that were not. Participants who did not experience a deviation event might not have spent as much time comparing the standard episodes, which resulted in lower recognition accuracy for variable details.

Deviation participants' extra time spent comparing episodes may not have resulted in higher recognition accuracy for fixed and new details compared to no-deviation participants for at least two reasons. First, recognition accuracy was relatively high for fixed details and it is possible that the presence of a deviation episode could not increase recognition accuracy any higher. Indeed, participants in repeated event studies that do not contain deviation episodes regularly recall fixed details very accurately [11]. Second, it might have been more difficult for deviation participants to compare new details, regardless of whether they were central or peripheral, between episodes as they only occurred once in each episode,

which is a lot less frequently than variable details that occurred in some form in every episode. If deviation participants did not spend much time comparing new details across episodes, this may not have resulted in more accurate memory for these details compared to no-deviation participants. Of course, it is also possible that our participants failed to notice the new details. Future research could further enhance the salience of new details by alerting participants to their presence before they are exposed to the standard episodes to see whether participants are able to “detect” these details when they occur. Past research has shown that content deviation effects are often associated with awareness of the deviations [14, 15].

Our fourth prediction, that late-deviation participants who experienced at least two standard episodes before the deviation episode should show better recognition accuracy for new details than early-deviation participants who had not experienced at least two standard episodes before the deviation episode, was not supported. Although there was a trend for new central details, it did not reach significance, and no other details showed a similar trend. It is possible that our participants had not moved from the schema-confirmation phase to the schema-deployment phase, which should have allowed them to devote more attention to script atypical details [31]. Our participants may have still been in the schema-confirmation phase and focused their attention on building the script rather than noticing new (atypical) details. This explanation is consistent with Rubínová et al. [15], who noted that, due to attentional demands, participants may not move into this second stage when new schemata are still developing. However, more research may be needed to establish evidence for schema-confirmation and deployment phases in adults.

We found that participants’ confidence in their memories for the target episode appeared relatively insensitive to the deviation episode: they were no more confident when a deviation episode was present than when it was not, and the timing and the similarity of the deviation episode had no effect on confidence. However, confidence varied by detail type: participants were more confident about variable details and new central details than fixed details and new peripheral details. This finding is somewhat surprising given that the fixed details stayed the same across standard episodes and, along with variable details, should have formed the basis of participants’ scripts. Fixed details also appeared to have the highest recognition accuracy.

This finding contributes to broader evidence that memory performance and metacognitive judgments can be partially dissociated, and that subjective confidence is not always sensitive to the same factors that affect memory performance and may instead rely on separate inferential or heuristic processes [e.g., 32].

The current findings have several implications. First, they suggest that deviation episodes may exert their strongest influence on the development and revision of event scripts, rather than on episodic memory performance more broadly. Although deviations consistently disrupted participants' predictions about future episodes, their effects on later recognition accuracy were limited. This pattern indicates that deviations may prompt observers to reconsider what typically occurs during a repeated event without

necessarily strengthening memory for all episode-specific details [see also 33]. The disruption produced by deviation episodes is consistent with event segmentation accounts suggesting that violations of expectations trigger updating of emerging event models and scripts, even when they do not necessarily translate into broad improvements in later episodic memory [34, 35]. In investigations of repeated victimisation, distinctive incidents may serve as natural anchors that help complainants separate episodes and attribute changing details more accurately. However, the limited benefits observed across other detail types suggest that investigators should not assume that unusual incidents produce uniformly superior memory for the surrounding event.

There are at least two limitations to the current study. First, participants experienced the standard and deviation episodes in one session, which may limit generalisability to real-world events that occur across days, weeks, and months. However, research examining the spacing of repeated episodes in which they occurred all in one day, across consecutive days or spaced over several days, showed that spacing had no effect on memory accuracy [7]. Second, participants had higher recognition accuracy for the moderate deviation episode than the high deviation episode. This may indicate that the moderate episode was more memorable than the high episode; however, there were no effects of similarity on participants' memory for the target event, which suggests that the higher recognition accuracy for the moderate deviation episode had little impact beyond the deviation episode. Instead, it might be that the details we asked participants about in the recognition test were slightly easier to answer for the moderate deviation episode than the high deviation episode, and do not actually represent differences in memory for these two episodes.

In conclusion, the findings showed our participants rapidly developed scripts for a repeated event and experiencing a deviation episode disrupted this process, particularly when the deviation episode had low similarity to the standard episodes. Low-similarity deviations produced greater disruption than high-similarity deviations. In contrast, deviations had only limited effects on participants' memory, improving recognition accuracy and reducing external intrusion errors for variable details but not enhancing memory for other details. These findings extend current understanding of script development and event cognition and have practical implications for contexts such as repeated victimisation, where distinctive incidents may help individuals differentiate between details that were experienced and those that were not without necessarily improving their accuracy in attributing details to particular episodes. Future research should examine the conditions under which deviations facilitate the transition from script formation to schema deployment and thereby enhance memory for atypical information.

Author statements

# Declarations

**Funding statement:** No funding was obtained.

**Author contributions statements:** All authors contributed to the main manuscript text and reviewed the manuscript.

**Disclosure statement:** The authors report there are no competing interests to declare.

**Declaration of generative AI use:** The authors report generative AI was not used in their research or preparation of this manuscript.

**Data availability statement:** The data are available through the OSF: https://osf.io/h5u7j/overview?view_only=8f78b723470943878c97d183f7c5650b

# References


1. Australian Bureau of Statistics. Sexual violence - victimisation. *Australian Bureau Statistics* (2021).
2. *Fair Work Act 2009* (Cth).
3. *PPP v. R* [2010] VSCA 110.
4. *R v. Baker* [2002] QCA 174.
5. Bell, B. E. & Loftus, E. F. Trivial persuasion in the courtroom: The power of (a few) minor details. *J. Pers. Soc. Psychol.* **56**, 669–679 (1989).
6. Nadel, L. & Simon, K. C. Perceived memory credibility: The role of details. *Proc. Natl Acad. Sci. USA* 121, e2416373121 (2024).
7. Danby, M. C., Sharman, S. J., van Golde, C., Paterson, H. M. & Watkins, R. The effects of episode spacing on adult's reports of a repeated event. *Memory* **31**, 879–889 (2023).
8. Dilevski, N., Paterson, H. M., Walker, S. A. & van Golde, C. Adult memory for specific instances of a repeated event: a preliminary review. *Psychiatry Psychol. Law*. **28**, 711–732 (2020).
9. Maclean, C. L., Coburn, P. I., Chong, K. & Connolly, D. L. Breaking script: Deviations and postevent information in adult memory for a repeated event. *Appl. Cogn. Psychol.* **32**, 474–486 (2018).
10. Theunissen, T. P. M., Meyer, T., Memon, A. & Weinsheimer, C. C. Adult eyewitness memory for single versus repeated traumatic events. *Appl. Cogn. Psychol.* **31**, 164–174 (2017).
11. Deck, S. L. & Paterson, H. M. Adults' ability to particularise an occurrence of a repeated event. *Appl. Cogn. Psychol.* **35**, 685–692 (2021).
12. Johnson, M. K., Hashtroudi, S. & Lindsay, D. S. Source monitoring. *Psychol. Bull.* **114**, 3–28 (1993).
13. Mitchell, K. J. & Johnson, M. K. Source monitoring: Attributing mental experiences. In The Oxford Handbook of Memory (eds Tulving, E. & Craik, F.) (2000). I. M.) 179–195 (Oxford University Press.
14. Rubínová, E., Blank, H., Ost, J. & Fitzgerald, R. J. Structured word-lists as a model of basic schemata: deviations from content and order in a repeated event paradigm. *Memory* **28**, 309–322 (2020).
15. Rubínová, E., Blank, H., Koppel, J. & Ost, J. Schema and deviation effects in remembering repeated unfamiliar stories. *Br. J. Psychol.* **112**, 180–206 (2021).
16. Connolly, D. A., Gordon, H. M., Woiwod, D. M. & Price, H. L. What children recall about a repeated event when one instance is different from the others. *Dev. Psychol.* **52**, 1038–1051 (2016).

17. Farrar, M. J. & Goodman, G. S. Developmental changes in event memory. *Child. Dev.* **63**, 173–187 (1992).
18. Schank, R. C. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People* (Cambridge University Press, 1982).
19. Schank, R. C. & Abelson, R. P. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures* (Lawrence Erlbaum, 1977).
20. Hudson, J. A. & Mayhew, E. M. Y. The development of memory for recurring events. In The Development of Memory in Infancy and Childhood (eds (eds Courage, M. L. & Cowan, N.) 2nd edn, 69–91 (Psychology, (2009).
21. Woiwod, D. M., Fitzgerald, R. J., Sheahan, C. L., Price, H. L. & Connolly, D. A. A meta-analysis of differences in children's reports of single and repeated events. *Law Hum. Behav.* **43**, 99–116 (2019).
22. Farrar, M. J. & Goodman, G. S. Developmental differences in the relation between scripts and episodic memory: Do they exist? In *Knowing and Remembering in Young Children* (eds Fivush, R. & Hudson, J. A.) 30–64Cambridge University Press, (1990).
23. Friedman, A. Framing pictures: The role of knowledge in automatized encoding and memory for gist. *J. Exp. Psychol. Gen.* **108**, 316–355 (1979).
24. Graesser, A. C., Gordon, S. E. & Sawyer, J. D. Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *J. Verbal Learn. Verbal Behav.* **18**, 319–332 (1979).
25. Graesser, A. C., Woll, S. B., Kowalski, D. J. & Smith, D. A. Memory for typical and atypical actions in scripted activities. *J. Exp. Psychol. Hum. Learn. Mem.* **6**, 503–515 (1980).
26. Neuschatz, J. S., Lampinen, J., Preston, E. L., Hawkins, E. R. & Toglia, M. P. The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Appl. Cogn. Psychol.* **16**, 687–708 (2002).
27. Smith, D. A. & Graesser, A. C. Memory for actions in scripted activities as a function of typicality, retention interval, and retrieval task. *Mem. Cognit*. **9**, 550–559 (1981).
28. McLindon, E. et al. Timelines of psychological, physical and sexual intimate partner violence among a nationally representative sample of Australian women. *Women's Health*. **21**, 17455057251329640 (2025).
29. Dianiska, R. E., Simpson, E. & Quas, J. A. Rapport building with adolescents to enhance reporting and disclosure. *J. Exp. Child. Psychol.* **238**, 105799 (2024).
30. Danby, M. C., Sharman, S. J., Brubacher, S. P. & Powell, M. B. The effects of episode similarity on children's reports of a repeated event. *Memory* **27**, 561–567 (2019).
31. Farrar, M. J. & Boyer-Pennington, M. E. Remembering specific episodes of a scripted event. *J. Exp. Child. Psychol.* **73**, 266–288 (1999).
32. Koriat, A. Oxford University Press,. The subjective confidence in one's knowledge and judgements: Some metatheoretical considerations. In *Foundations of Metacognition* (eds Beran, M. J., Brandl, J.

L., Perner, J. & Proust, J.) 213–233 (2012).

33. Kontogianni, F. et al. Facilitating recall and particularisation of repeated events in adults using a multi-method interviewing format. *Memory* **29**, 471–485 (2021).
34. Nolden, S., Turan, G., Güler, B. & Günseli, E. Prediction error and event segmentation in episodic memory. *Neurosci. Biobehav Rev.* **157**, 105533 (2024).
35. Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S. & Reynolds, J. R. Event perception: A mind-brain perspective. *Psychol. Bull.* **133**, 273–293 (2007).

## Footnotes

1. A Greenhouse-Geisser correction was applied due to assumption of sphericity being violated.

## Figures

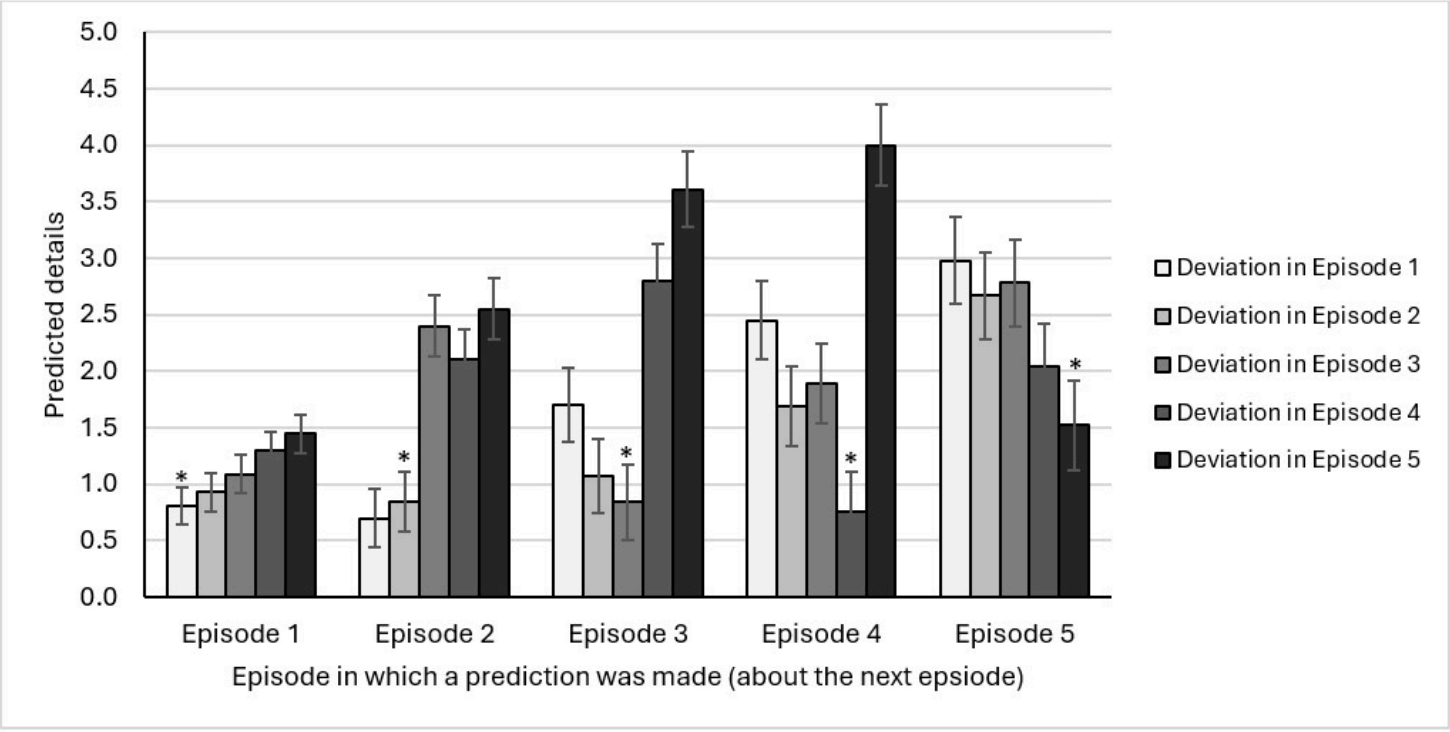



**Figure 1**

Participants' predicted number of details across episodes. *Episode in which the deviation occurred.