Cognition 249 (2024) 105833

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Event representation at the scale of ordinary experience

Sami R. Yousif [a,*], Sarah Hye-yeon Lee [b], Brynn E. Sherman [a], Anna Papafragou [b]

[a] University of Pennsylvania, Department of Psychology, United States of America
[b] University of Pennsylvania, Department of Linguistics, United States of America

ARTICLE INFO

Keywords:
Event representation
Episodic memory
Temporal memory

ABSTRACT

Weeks are divided into weekdays and weekends; years into semesters and seasons; lives into stages like childhood, adulthood, and adolescence. How does the structure of experience shape memory? Though much work has examined event representation in human cognition, little work has explored event representation at the scale of ordinary experience. Here, we use shared experiences — in the form of popular television shows — to explore how memories are shaped by event structure at a large scale. We find that memories for events in these shows exhibit several hallmarks of event cognition. Namely, we find that memories are organized with respect to their event structure (boundaries), and that beginnings and endings are better remembered at multiple levels of the event hierarchy simultaneously. These patterns seem to be partially, but not fully, explained by the perceived story-relevance of events. Lastly, using a longitudinal design, we also show how event representations evolve over periods of several months. These results offer an understanding of event cognition at the scale of ordinary human lives.

Experiences are complex and varied, but our memories of them are structured in the form of discrete events. We remember our high school years as distinct from our college years and those years as distinct from everything after. We remember our summer months as distinct from the grind of the semester. We remember our weekdays as distinct from our weekends.

A significant body of work in psychology and cognitive science has been interested in the nature of event representation — how it is that we carve *continuous* experience up into *discrete* 'units' (for reviews, see Radvansky & Zacks, 2011; Shipley & Zacks, 2008; Yates, Sherman, & Yousif, 2023; Zacks, 2020). Most of this work, unsurprisingly, has focused on events that can be easily studied in the laboratory: events with a simple linear structure, typically lasting seconds, minutes, or hours, at most. Yet our lives unfold not on the timescale of moments or minutes, but weeks and months and years. And any one 'moment' may be part of a representation at multiple scales (e.g., on a Friday afternoon, on the 9th of October, in the year 2020). Moreover, our lives are not one single narrative: There are overlapping plots, a revolving door of characters, and unexpected twists (see Conway, 1996). What does event representation look like at this scale — for events that share the 'texture' of ordinary experience?

1. Event representation at shorter vs. longer timescales

In everyday language as well as in cognitive science, an "event" may refer to many different things (see Yates et al., 2023). Some work emphasizes *event boundaries* (see, e.g., Radvansky & Copeland, 2006; Radvansky, Krawietz, & Tamplin, 2011; Radvansky, Tamplin, & Krawietz, 2010; see also Radvansky, 2012). Other work subtly manipulates context (e.g., a change from an orange to a purple background; DuBrow & Davachi, 2013; DuBrow & Davachi, 2016; Ezzyat & Davachi, 2014; Heusser, Ezzyat, Shiff, & Davachi, 2018). Yet other work relies on more naturalistic events, in the form of written stories (Copeland, Radvansky, & Goodwin, 2009; Doolen & Radvansky, 2021, 2022; Speer, Zacks, & Reynolds, 2007) or video media (e.g., Baldassano et al., 2017; Yates et al., 2022; Zacks, Speer, Swallow, & Maley, 2010). While these studies certainly reveal something about the nature of event processing, they do not fully reflect the richness of ordinary experience (but see Teigen, Böhm, Bruckmüller, Hegarty, & Luminet, 2017; Rouhani et al., 2023).

Current theories of event cognition point to two influential ideas about event representation. The first is the notion that event *boundaries* influence memory in a variety of ways. For instance, boundaries influence associative memory: The temporal order of items across events is more easily confused than temporal order within events, suggesting that temporal memory is organized around event structure to some extent

* Corresponding author.
E-mail address: sryousif@sas.upenn.edu (S.R. Yousif).

https://doi.org/10.1016/j.cognition.2024.105833
Received 26 September 2023; Received in revised form 15 May 2024; Accepted 22 May 2024
Available online 3 June 2024

(DuBrow & Davachi, 2013; DuBrow & Davachi, 2016; Ezzyat & Davachi, 2011; Heusser et al., 2018; Pu, Kong, Ranganath, & Melloni, 2022). This is further supported by evidence that there is greater neural representational similarity for items within vs. across events (DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014; Ezzyat & Davachi, 2021). Boundaries also influence memory for items that occurred on either side of the boundary. Sometimes, memory is disrupted when there is a shift in context (e.g., when walking through a doorway; Radvansky, 2012). Other times, event memory may be enhanced right at event boundaries (e.g., Swallow, Zacks, & Abrams, 2009), perhaps reflecting increased attention (Pradhan & Kumar, 2021; Radvansky & Zacks, 2017; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Boundaries also influence the perception of time. For instance, events with boundaries are perceived as shorter in duration than continuous events (Liverence & Scholl, 2012; Yousif & Scholl, 2019; Sherman, DuBrow, Winawer, & Davachi, 2023). This is also reflected in how memories are remembered relative to one another: Items spanning event boundaries are remembered as farther apart in time than items contained within a single event (Bangert, Kurby, Hughes, & Carrasco, 2020; Clewett, Gasser, & Davachi, 2020; Ezzyat & Davachi, 2014; Faber & Gennari, 2017). Although these findings are well-established for short-term event memory, it is unclear whether these same effects will manifest on longer timescales and for highly structured, hierarchical events.

A second key idea in the study of event cognition is that event beginnings and endings – *qua* event boundaries – might have distinct signatures from other timepoints such as event middles (see, e.g., Ji & Papafragou, 2020a, 2020b; Ji & Papafragou, 2022; Teigen et al., 2017; Yates et al., 2023). Some work has emphasized the importance of beginnings (e.g., Teigen et al., 2017) whereas other work has emphasized the importance of endings (Ji & Papafragou, 2020a, 2020b; Ji & Papafragou, 2022; see also work on 'primacy' vs. 'recency' and 'serial position curves', e.g., Doolen & Radvansky, 2021; Kahana, Diamond, & Aka, 2022; McCrary Jr & Hunter, 1953; Murdock Jr, 1962; Roediger III & Crowder, 1976; Sehulster, 1989). However, we currently lack evidence of how beginnings and endings are represented within longer, more complex events containing multiple nested units. This question is about more than timescale: Complex, extended events might fade or be reorganized in memory over time. Little is known about such reorganization.

A prominent theoretical suggestion is that more complex event representations are realized via *situation models*. For instance, both Kurby and Zacks (2012) and Zacks, Speer, and Reynolds (2009) examined narrative comprehension of scenes from the book *One Boy's Day*, which included analyzing people's reactions to the narrative as they read it. The key takeaway from this work is that event segmentation is the result of changes in situational features such as characters and locations. Relatedly, work in the autobiographical memory literature has examined how multifaceted memories are shaped and reshaped over time (see Bruner, 1994; Conway & Pleydell-Pearce, 2000; Conway & Rubin, 2019; Yang, Deffler, & Marsh, 2022). Some work, for instance, has examined how memories are influenced by the structure of the calendar year. The *calendar effect* (or *end point effect*) describes the tendency for people to freely recall events from the beginning and ending of the calendar year more than events from other periods (Kurbat, Shevell, & Rips, 1998; Thomsen & Berntsen, 2005). These findings suggest that event structure, independently from event content, may indeed influence human memory.

2. Current study

Here we probed events at larger scales than those typically studied in the laboratory, reaching the scale of ordinary experience both in their duration (months, years) and their complexity (insofar as they contain multiple nested units). To do that, we tested memories for shared experiences — popular television shows with extended narratives — that span weeks or months or years. We picked three different shows, *Game of Thrones* (Experiments 1a-c), *Stranger Things* (Experiment 2), and *The Last of Us* (Experiments 3a-c), each of which had unique properties. In this way, the results here reflect general properties of event representation, rather than properties of one or two specific narratives. We chose to study television shows because we viewed them as a useful 'case study' for understanding event representation in the real world (for a direct comparison, see Yang et al., 2022). Our work draws inspiration from prior work on event cognition and autobiographical memory to understand not just what events people recall from structured narratives, but also the shape those narratives take on a mental timeline as a result of such structure.

At the scale of ordinary experience, there are many 'natural' event boundaries — the beginning of a new school year, the birth of a child, a Bat Mitzvah. These boundaries may segment experience into distinct events, or 'units'. Here, we studied how memories of television shows are organized with respect to their overarching event structure using the obvious, hierarchically organized units of *episodes* and *seasons*. We investigated two key aspects of event representation: the structure of memories with respect to boundaries (e.g., within vs. between episodes, or seasons), and the relative memory for beginnings and endings across levels of the event hierarchy. The choice to analyze seemingly arbitrary breakpoints of the show might appear surprising: after all, when we read a novel, we are unlikely to remember the contours of its paragraphs and chapters in any detail but very likely to remember subparts of the plot. However, our choice is motivated by the fact that the surface-level organization of the shows into episodes and seasons are likely related to natural plot breakpoints. Writers naturally place story-relevant events at beginnings and endings of units on purpose (and it is this expectation that makes cliffhangers striking). Similarly, the opposite might be true: proximity to beginnings or endings might influence perceived meaningfulness. Studies on autobiographical memory have shown that people automatically recall events at the beginning and end of the calendar year more often than events in the middle, for instance (Kurbat et al., 1998; Thomsen & Berntsen, 2005). Events that are more likely to be remembered may also become more likely to seem integral in a larger story. In other words: It is possible that arbitrary boundaries, by virtue of influencing what is remembered, also influence perceived meaningfulness. We address these two possibilities explicitly in the studies that follow.

Another possible limitation of this approach is the fact that people may have experienced the shows we study differently: Some may have watched each episode with regularity as they were aired, while others may have watched dozens of episodes in a single weekend. For this reason, the narrative may or may not be separated by large gaps and many life experiences. If we did observe event related effects, one may rightfully wonder whether these effects reflect the structure of the shows and their stories, or the simple fact that there were likely at least some temporal gaps in the narrative, per the nature of how television shows are viewed. However, it is not obvious that real world experiences are all that different from TV shows in terms of how they are structured (for an empirical perspective, see Yang et al., 2022). The natural boundaries in our lives — for instance, the transition from high school into college, or from college into the real world — often do coincide with meaningful structures that we use to understand our experience. Moving to a new city may mark the end of a relationship, or, as well, the end of a relationship might be a cause for moving to a new city. In television shows, narrative events may be purposefully structured within boundaries. But in life, naturally occurring, unscripted events are often structured in much the same way, even if by accident. Therefore, these results may be an apt, even if imperfect, case study of event representation more broadly.

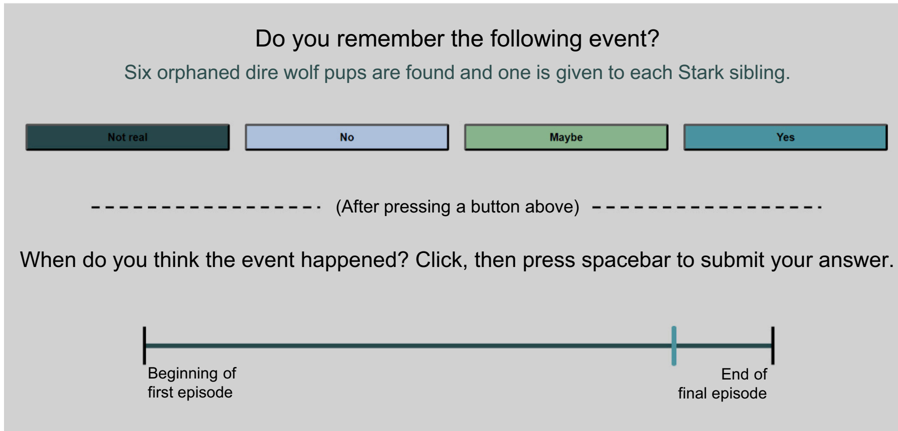

Fig. 1. A depiction of the task (here, the sample event is from *Game of Thrones*). For each event, participants were first asked the generic memory question and could respond in one of four ways ("Not real", "No", "Maybe", or "Yes"). After they responded, the four buttons disappeared, and a prompt to place the event on a timeline appeared.

3. Experiment 1: Game of Thrones

In Experiment 1, participants were asked to place events from *Game of Thrones* on a simple timeline, ranging from the first episode to the last. We explored how memories were distorted relative to the event structure (i.e., seasons, episodes) of the narrative. We chose to study *Game of Thrones* because it has a single, overarching story that was shown on television over the course of nearly a decade. The story consists of 73 episodes over 8 seasons, with many dozens of characters and settings. Moreover, as the show has concluded, we can assess memory for a fully bounded, complete experience.

3.1. Experiment 1a — Simple timeline

3.1.1. Methods

All aspects of the procedure and design (for all experiments) were pre-registered prior to data collection. Those pre-registrations, as well as raw data and materials, can be found at our OSF page: https://osf.io/syw84/https://osf.io/syw84/

3.1.1.1. Participants. Participants were recruited via the Prolific platform. They were encouraged not to participate unless they had seen the entire show. There was a final sample of 50 participants, after exclusions and replacement. Per our pre-registered criteria, participants were excluded only if (a) they failed to correctly identify the fake attention-check items (indicated by an average score of 1 or higher if responses are scored as follows — Yes: 2, Maybe: 1, No: 0, Not real: −1); (b) they identified >20% of the events as 'fake', or (c) the overall correlation between participants timeline responses and the serial position of the items was <0.30. Eight participants from Experiment 1a were excluded based on these criteria.

3.1.1.2. Procedure. The experiment was administered online via a web-based interface using custom JavaScript code. The task itself consisted of a simple memory question and then a more specific timeline response. Participants were probed about a specific event (order fully randomized across participants) from the show and then asked to indicate whether they remembered that event (see Fig. 1). There were four possible options: Yes, maybe, no, or 'not real'. Participants clicked on an icon to indicate their response. Regardless of their response to the memory question, they were then asked to place that event on a timeline from the first episode to the last. For Experiment 1a, the timeline was made up of a dark teal (#2D4F4D) line with vertical black lines at the two endpoints. The timeline was 780 pixels wide, and so responses were coded as ranging from −390 to 390 along the x-axis.[1] Participants could point and click on the timeline and then drag and drop a response marker until they were satisfied. They pressed spacebar to submit their responses, at which point another item would be tested after a short delay. There was no time limit on participants' responses. Participants responded in this way until they had seen all of the events.

After participants responded to the final event, they were redirected to a brief survey. The survey varied slightly from experiment to experiment, but in general we asked (1) How many times they'd seen the show; (2) Over what time span they had seen the show; (3) When they had last watched the show; (4) Whether they'd watched some parts more than others; and (5) How well they thought they did overall. Data from these surveys (for all experiments) are included along with the raw data on our OSF page. The survey data lack the granularity necessary to conduct detailed analyses, but they do give a sense of the participant population. One thing to note is that there was a considerable heterogeneity in the data set with respect to whether people watched the shows over a long vs. short period of time, whether they had watched it multiple times, etc.

We tested 2 items per episode. Because *Game of Thrones* had 73 total episodes, this meant that there were 146 real events and 4 fake events (for a total of 150). For this experiment and all subsequent experiments, the events were chosen as systematically as possible to avoid bias. We pulled these events from Wikipedia summaries of each episode. We modified the language as little as possible and did so only when modifications were necessary to improve clarity or provide context (fewer than 10% of the events across all experiments included modified language.) As much as possible, these events referred to discrete moments rather than periods of time (e.g., "Arya arrives at Winterfell" rather than "Arya travels to Winterfell", as the latter may span multiple episodes). We avoided referring to attitudes/goals/emotions as much as possible. The exact language of the events that we used can be found along with

[1] For a subset of participants in Experiment 1a, the timeline response for trials where a participant responded 'fake' (regardless of whether the event was real or not) were erroneously recorded as −400. This issue affected 60 trials in total (26 trials in which the events were real); these trials were excluded from all timeline-based analyses.

other materials on our OSF page.

3.1.2. Results and discussion

As a general tool for assessing memory strength, we first computed a ‘subjective memory score’ for each event. Specifically, we coded people’s responses to the initial memory question in the following way: Yes: 2, Maybe: 1, No: 0, Not real: −1. Thus, events with an average subjective memory score of between 1.5 and 2 indicate that the majority of participants remembered that event with certainty. Overall, participants reported remembering the vast majority of the 146 events that were tested. The average subjective memory score for each real event was 1.6 (SD = 0.28), indicating that for most events participants were somewhere between “Maybe” and “Yes”; the average for the fake events was −0.67 (SD = 0.39). Only three events had an average score below 1; one event had an average score below 0.5.

However, our primary measure of interest was participants’ temporal memory errors. A summary timeline reflecting the median responses for each event can be seen in Fig. 3A. As a measure of overall temporal source memory, we computed the Spearman rank correlation between the placement of an individual event on the timeline and its actual position within the event sequence. Importantly, the rank order nature of this correlation captures the extent to which participants are sensitive to the *relative order* of the events, rather than whether their placements of the events reflected the true absolute position of the event within the entire series. In this way, a participant could dramatically misuse the timeline (e.g., by only using half of it) and the data would still be fully interpretable. This, and other key analyses, therefore sidesteps any concerns one may have about participants misusing or misjudging the timeline. Across participants, the average correlation was $rho = 0.75$ ($SD = 0.11$; $p < .001$; See Supplemental Fig. S1). In other words, participants have fairly robust temporal memory overall.

For our main analyses, we asked how memories of television shows are *organized* with respect to their hierarchically organized units of *episodes* and *seasons*. We investigated two key aspects of event representation: (A) The structure of memories with respect to boundaries (e.g., within vs. between episodes, or seasons), and (B) The relative memory for beginnings and endings across levels of the event hierarchy.

3.1.2.1. Boundaries. We evaluated the effects of boundaries in two primary ways. First, we asked whether units (e.g., seasons) were ‘temporally compressed’ relative to one another, such that there were clear ‘gaps’ in the timeline between units. For instance, if seasons were temporally separated (see, e.g., Clewett et al., 2020), events *within* a season should be remembered as clumped together (as in Fig. 2B). Conversely, we might expect that after long temporal delays, this effect reverses — that, as memories become noisier, the extent of each season

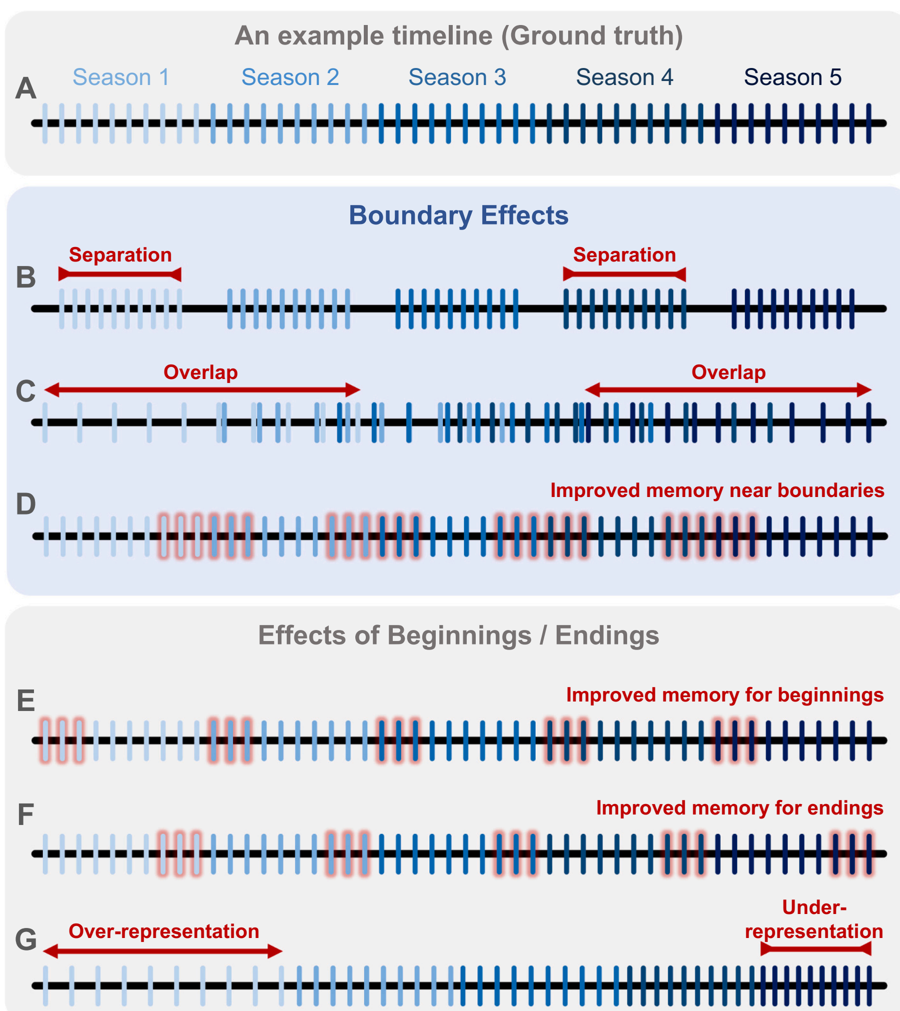

Fig. 2. Possible patterns of results for the timeline analyses, here depicted for a hypothetical show with 5 seasons of 10 events each. Each vertical bar represents an event. Each season is depicted in a distinct color. The first panel (A) depicts a hypothetical baseline; all other results are depicted relative to this baseline. Red highlighting (D, E, and F) depicts improved memory. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

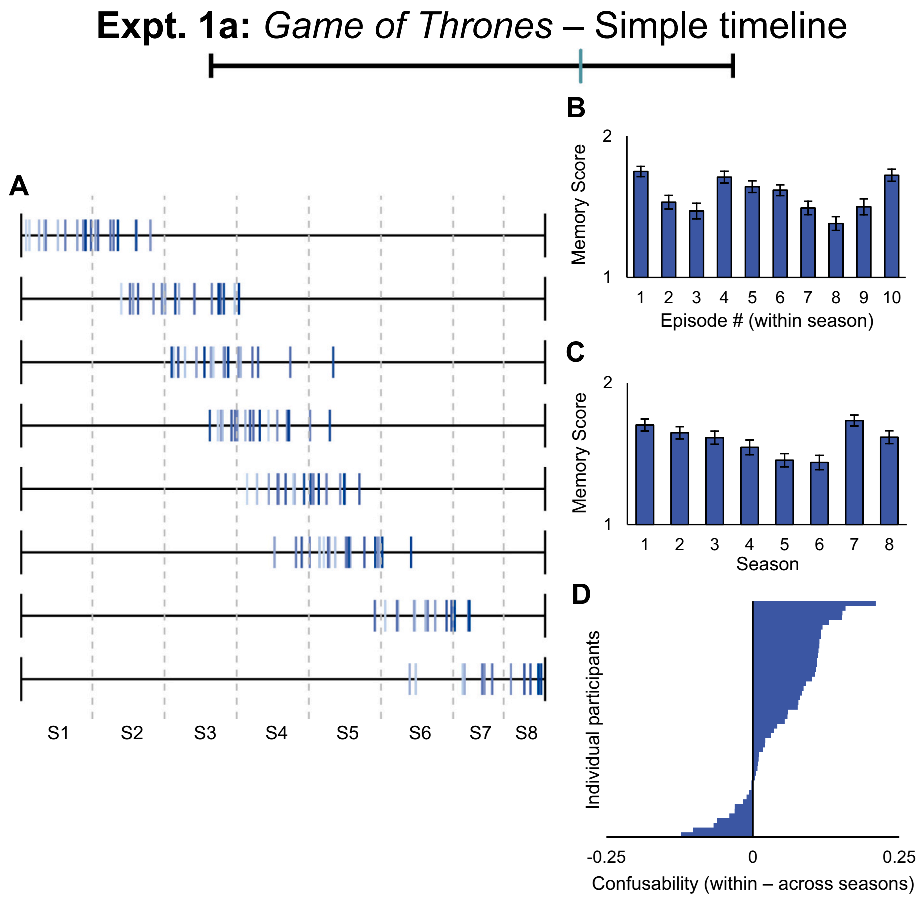

Fig. 3. Results of Experiment 1a. (A) Median timeline placement by season (from first [top] to last [bottom]) and episode (events in the earlier episodes in each season are depicted as greyer in color; the later episodes are depicted in a bluer color). The vertical grey dotted lines represent the season boundaries. Note that the medians are plotted here for visualization purposes; most analyses were conducted on the individual subject data. (B) The mean subjective memory score for each episode, aggregated across seasons. (C) The mean subjective memory score for each season. (D) The 'confusability difference score' (see *Methods*; Fig. 4) for each participant. Error bars represent +/− 1 SE.

on the mental timeline expands and the seasons begin to bleed into one another (as in Fig. 2C).

To assess temporal separation, we calculated an 'overlap score' which quantified the extent to which the representations of units overlapped with one another. We measured — for each event — whether the placement of that event is *before* the placement of an event from an earlier unit (e.g., season, episode) or *after* the placement of an event from a later unit. We then summed the total number of overlaps and divided by the total number of events. Because items can overlap from other items across multiple seasons at once, these percentages can be >100%. Note that if there was no overlap across units (i.e., if units were perfectly temporally segregated, as in Fig. 2B), the overlap score would be 0.

Fig. 3A suggests that seasons were *not* temporally compressed. Indeed, the opposite seemed to be true: The seasons appeared to bleed over into one another. This metric yielded an 'overlap score' of 142%, confirming what is evident in the figure. (Note that if the seasons were represented as non-overlapping, there should be 0% overlap across the seasons. Also note that unlike most of the analyses reported here, this analysis was conducted on the group-aggregated data in order to reduce noise; however, we observed similar, albeit noisier, patterns when conducting this analysis separately for each individual participant.)

The critical question, however, is whether temporal memories degrade with respect to the event structure. We can quantify how event structure influences the organization of memory by asking whether participants made more temporal memory errors (i.e., confusions in temporal order) for items within vs. across seasons. One possibility is that seasons would be represented as distinct from one another; if so, there should be more confusions *within* a season than *across* seasons (controlling for temporal distance). This is akin to classic categorization effects (see Murphy, 2004). Alternatively, there could be more confusions *across* seasons rather than within. There is precedent for such a finding within the event cognition literature (see e.g., DuBrow & Davachi, 2013; DuBrow & Davachi, 2016; Ezzyat & Davachi, 2011). However, these results relied on events that were ordered arbitrarily, unlike the present events that were part of a larger narrative structure.

For these analyses, we calculated 'confusability difference scores' within vs. across seasons (see Fig. 4). To do this, we took each event and calculated the number of events within a certain window that were out of order with that event (i.e., a confusion). We then quantified the proportion of confusions with items *within* the same unit (season) versus *across* unit boundaries. From this, we can calculate a difference score of 'confusability' within vs. across seasons. A difference in confusability within vs. across season boundaries would indicate that the boundaries influenced the encoding of events relative to one another. Confusability difference scores were calculated separately for each participant.

Here, the 'window' over which this analysis was conducted was +/− 5 episodes (±10 events). Because of this, we did not calculate 'confusability' scores for first five and final five episodes. Note that this window was slightly different from some of the subsequent experiments. The results of the analyses are not dependent on the size of these windows. Further, we note that the number of confusions were always normed to the total number of possible within vs. across event confusions (i.e., we calculated the proportion of times an episode was confused with another

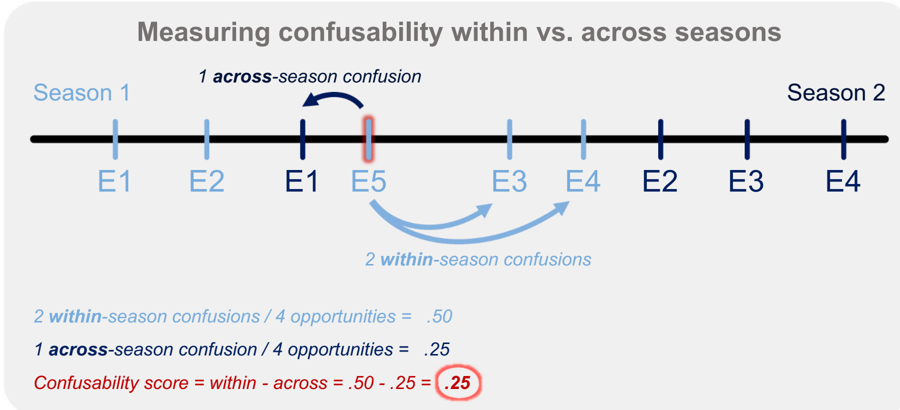

Fig. 4. A visual explanation of the 'confusability' metric. This is a made-up example with two seasons of five episodes/events. The analysis is conducted separately for each event. Here, it is being conducted on Event #5 in Season 1. The placement of this event is confused with two events from the same season but only one from a different season, resulting in a positive 'confusability' score.

episode in its season, relative to the total number of episodes from that season within the 5-episode window), so our results cannot be explained by, e.g., within-event confusions being statistically more likely.

Across participants, the average confusability difference score was 5% (SD = 7.1%), such that participants made on average 5% more confusions within vs. across seasons; this was reliably greater than would be expected by chance ($t(49) = 5.02$, $p < .001$, $d = 0.71$; Fig. 3D). In other words, despite the fact that seasons were represented as overlapping with one another in time, temporal order memory was still shaped by the structure of the seasons.

Note that these effects are consistent with classic categorization effects (see Murphy, 2004) but the opposite of what has been observed for events studied on shorter timescales (for which items across boundaries are more confusable with one another than items within a single event; e.g., DuBrow & Davachi, 2013).

Why might this be the case? We see two possibilities. One possibility is that the structure of events in the current study is different from that of prior work. In the current study, the relevant events, or periods, are not arbitrary. Participants were not switching between one random context and another. Each season of the show has a meaningful structure, and the ordering of events across seasons is perhaps more salient than the ordering of events within seasons. Intuitively, this makes sense: The mind may not care about the ordering of events across units if those units are arbitrary but should care if they are a part of some higher-order structure (as would be the case for a television show). Indeed, recent work found that when events comprise meaningful structure (i.e., all items in Event 1 were animals and all items in Event 2 places), people exhibited better temporal memory for items from different events (Wen & Egner, 2022). A second possibility is that there are genuinely opposing effects at different timescales. Indeed, there are several other examples of event-related effects which go in opposite directions at different timescales (see, e.g., McNerney, Goodwin, & Radvansky, 2011; see Yates et al., 2023 for a discussion of how event structure has opposing effects on perceived time depending on the timescale). It seems entirely plausible, then, that there are other qualitative differences between short- and long-term event representation.

To better understand the relation between memory and these boundary effects, we ran two exploratory (not pre-registered) analyses. First, to understand how participants overall memory (as measured by whether or not they remembered an individual event) was related to the temporal source memory errors they made, we asked whether the subjective memory score for each event was related to the average deviation in the placement of that event on the timeline (absolute difference of the mean rank of the event across participants, relative to the true rank position of that event on the timeline). We observed a reliable, negative correlation, $r = -0.34$, $p < .001$, suggesting that events that were more reliably remembered across participants were also remembered more precisely. Second, to more directly get at the question of how memory relates to the observed boundary effects (greater confusability within vs. across seasons), we related each participant's Spearman rank correlation (a measure of their overall temporal memory) with their confusability difference score. We observed a positive correlation between these metrics ($r = 0.29$, $p = .043$), suggesting that participants with more precise temporal memories also made relatively fewer across, relative to within, season confusions. In other words, participants with better memories overall also more strongly segmented with respect to season boundaries.

3.1.2.2. Beginnings and endings. Next, we investigate whether and how the beginnings and endings of events are better remembered. If beginnings and endings are special at every level of an event hierarchy, we may expect better event memory for items at the beginning (see Fig. 2E) and ending (see Fig. 2F) of each season, as well as at the beginning and ending of an entire show (see, e.g., Pu et al., 2022; Wen & Egner, 2022).

To quantify whether memory for the beginnings and endings of each season was enhanced, we computed the average subjective memory score for each episode (averaged across seasons) and assessed whether the subjective memory score for the first and last episodes were greater than the intermediate episodes. There was a main effect of episode on memory ($F(9,441) = 27.6$, $p < .001$, $\eta_p^2 = 0.36$; Fig. 3B), with Episodes 1 and 10 having the highest average accuracy. To further probe this pattern, we ran pairwise tests to assess whether Episodes 1 and 10 respectively were remembered significantly better than the other episodes. Episode 1 had higher accuracy than all other episodes ($ps < 0.002$, all of which survived Bonferroni correction at $\alpha = 0.006$), except 4 and 10; Episode 10 had higher accuracy than all other episodes ($ps < 0.004$, all of which survived Bonferroni correction at $\alpha = 0.006$), except 1 and 4.

Next, to probe whether memory for the beginnings and episodes of the entire show was enhanced, we computed the average subjective memory score for each season (averaged across episodes) and assessed whether the subjective memory score for the first and last seasons were greater than the intermediate episodes. There was also a main effect of season on memory ($F(7,343) = 17.8$, $p < .001$, $\eta_p^2 = 0.27$; Fig. 3C). Across seasons, accuracy decreased from the first season until the sixth, at which point there was a large jump in accuracy. Season 7 was the most well remembered, with reported memory significantly higher than all other seasons as measured via pairwise tests to all other seasons ($ps < 0.04$, five of which survived Bonferroni correction at $\alpha = 0.007$) except Season 1 ($p = .48$). Reported memory was higher for Season 1 than all other Seasons except Season 7 ($ps < 0.06$, four of which survived Bonferroni correction at $\alpha = 0.007$). This is largely consistent with prior work on long-term memory of events (e.g., Doolen & Radvansky, 2021, 2022; Roediger III & Crowder, 1976) and the long-established notion of a 'serial position curve' (see, e.g., McCrary Jr & Hunter, 1953).

We also assessed whether certain events (i.e., events from the beginning and ending of the show) were remembered in a consistent way across participants. We thus assessed the 'consistency' of individual event memories by computing the standard deviation in the rank order of participants' placements on the timeline. Note that unlike most of the analyses reported here, this analysis is computed on the item level (i.e., over *events* rather than over participants). Of the twenty events remembered with the greatest temporal consistency across participants (we chose the number 20 because it corresponds to the number of events we tested from each normal-length season of the show), 18 were from first or the final two seasons. To assess whether this is significantly higher than would be expected by chance, we ran a permutation test, in which we generated a null distribution by scrambling the season-consistency score mappings (x1,000) and recomputing how many of the 20 most precisely remembered events were from the first or final two seasons. Assessing the proportion of permutations in which the observed effect (18/20) was greater than the null yielded a permutation $p < .001$, suggesting that this pattern of results was highly unlikely to be expected by chance. (Note that *Game of Thrones*' final two seasons were shorter than other seasons and thus distinct from them; that is why they are being combined for these analyses.)

Second, we asked whether beginnings (and endings) were relatively *over- or under-represented* in the sense of taking up more (or less) space in a show's mental timeline than they should (see Fig. 2G; see, e.g., Teigen et al., 2017). To compare the relative over- and under-estimation of the length of each unit (i.e., season), we also conducted an analysis measuring the distance between the first and last events for each unit. We computed this metric separately for each participant. To minimize noise in Experiments 1 and 2, we averaged the position for the two events of each episode (i.e., for *Game of Thrones* Season 1, we computed the distance from the average placement of the two events from Season 1, Episode 10 to the average placement of the two events from Season 1, Episode 1. We then divided this 'representational distance' by the actual temporal distance between the first and last events of that season to compute an over-representation metric; values >1 indicate that the unit was over-represented, relative to the true length of the season, whereas values <1 indicate that the unit was under-represented. Note also that values here can be negative; if a participant misremembered the order of events within a unit and placed the final events occurring before the first event, then the 'representational distance' would be negative. Via this metric, Seasons 1, 7, and 8 were reliably over-represented across participants ($ps < 0.005$). Season 8, on average, occupied nearly three times as much 'mental real estate' as it should have; this was followed by Season 7, which occupied nearly two times as much, and then Season 1, which occupied 1.4× the space it should have. Seasons 3 and 4, in contrast, were reliably under-represented ($ps < 0.002$), with Season 4 being the least represented. All of these over- and under-representation contrasts survived Bonferroni correction at $\alpha = 0.006$.

Collectively, these analyses provide evidence that beginnings and endings are prioritized in memory. Individuals best remembered events that occurred earlier and later both *within* seasons and *across the entire show*, and earlier and later seasons were overrepresented on people's mental timelines. Additionally, in a complementary analysis, we demonstrated that individual events from earlier and later seasons were remembered with the greatest temporal consistency across people.

3.2. Experiment 1b — Demarcated timeline

A potential concern for Experiment 1a was that people failed to use the timeline correctly by misjudging where season boundaries would have been located. In Experiment 1b, we modified the timeline to address this concern.

3.2.1. Methods

3.2.1.1. Participants. A separate group of participants were recruited via Prolific for Experiment 1b. Nine participants were excluded based on our pre-registered criteria; as in Experiment 1a, the final sample included 50 participants after exclusions and replacement.

3.2.1.2. Procedure. The procedure was identical to Experiment 1a. However, each individual season of the show was demarcated on the timeline; vertical bars separated the seasons, and words beneath the timeline clearly labeled each season (see Fig. 5).

3.2.2. Results and discussion

We observed all of the same patterns as in the original version of the task in Experiment 1a (see Fig. 5; see full analyses in the *Supplementary Materials* on the OSF page). Additionally, the median placement of each of the 146 events was almost perfectly correlated across the two versions ($rho = 0.98$, $p < .001$).

These results suggest that the prior data cannot be explained by a failure to understand the timeline: Participants consistently misremembered some events from Season 1 as occurring well within the boundaries of Season 2 and even Season 3, even when there were explicit labels demarcating each season. This pattern is highly robust in the data. Recall that what is plotted in Fig. 5 is *median* position, meaning that if an event from Season 1 is plotted within Season 2, at least 50% of participants must have placed that event in the wrong season. Many events were even placed multiple seasons ahead/behind. If participants were correctly recalling when events occurred, but failing to use the timeline, we should have observed far fewer errors of this sort in Experiment 1b compared to Experiment 1a. Thus, we are very confident that the results observed in Experiment 1a (and our other experiments) reflect genuine distortions of human memory rather than misuse of the timeline.

3.3. Experiment 1c — Story relevance

Experiments 1a and 1b provide evidence that participants' memories are shaped by the temporal structure of the series. To what extent are these effects driven by the narrative structure of the show? For example, are beginnings and endings of seasons (and the series) better remembered because important events just happen to occur at the beginning and ending of each season (e.g., as part of a narrative choice made by the writers)? Or are enhanced memories for beginnings and endings unrelated to the show's narrative structure? In Experiment 1c, we address this question by asking an independent sample of participants to assess the importance of each event to the overall narrative.

3.3.1. Methods

3.3.1.1. Participants. A separate group of participants were recruited via Prolific for Experiment 1c. One participant was excluded based on our pre-registered criteria; the final sample included 30 participants after exclusions and replacement.

3.3.1.2. Procedure. In Experiment 1c, rather than a timeline, participants were asked to indicate how story-relevant each event was, on a scale of 1 (not relevant at all) to 5 (extremely relevant).[2] Participants

[2] One might wonder whether the relevance question truly captures the causal structure of the narrative. We ran an additional version of the study, in which we instead asked participants an explicit counterfactual question as a measure of causal dependence, namely, "how much the story would have changed if that event had not occurred". These ratings were highly correlated to the relevance ratings reported in Experiment 1c and yielded qualitatively identical patterns to what is reported here. We have included these data on our OSF page.

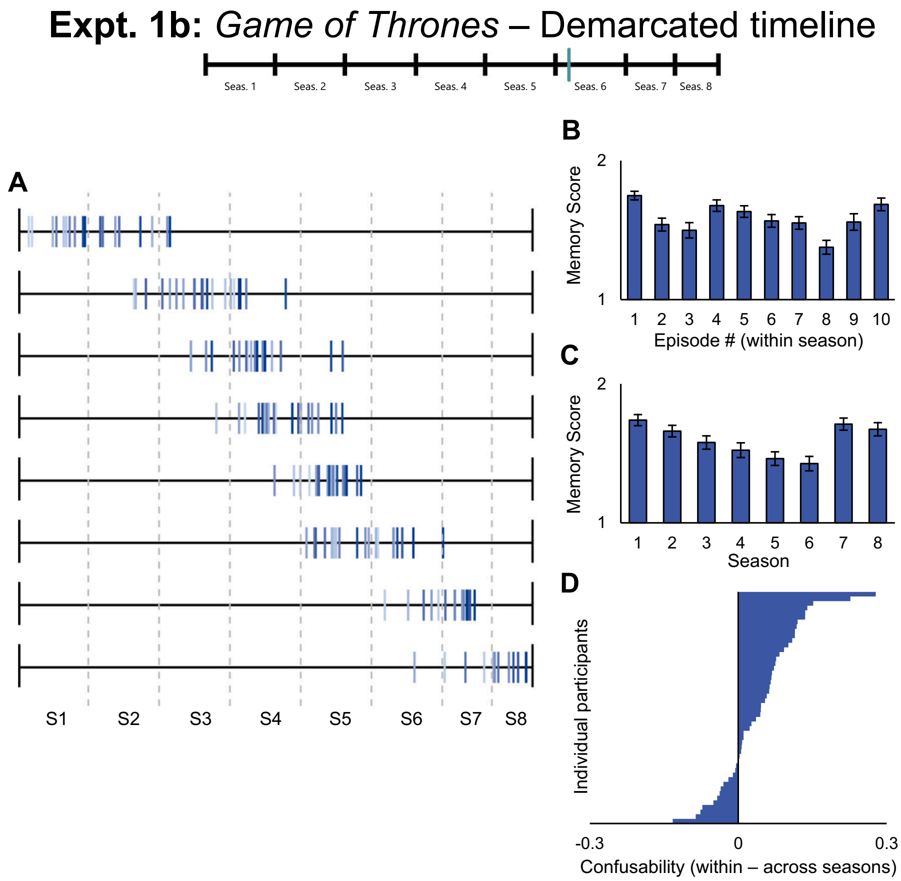

Fig. 5. Results of Experiment 1b. (A) Median timeline placement by season (each horizontal line represents a distinct season, from first [top] to last [bottom]) and episode (the earlier episodes in each season are depicted as greyer in color; the later episodes are depicted in a bluer color). (B) The mean subjective memory score for each episode, aggregated across seasons. (C) The mean subjective memory score for each season. (D) The 'confusability difference score' (see *Methods*; Fig. 4) for each participant. Error bars represent +/− 1 SE.

indicated their responses by clicking on a radio button and then pressing a "Submit" button.

3.3.2. Results and discussion

In a first set of analyses, we assessed story relevance judgments by episode. There was a main effect of episode ($F(9,261) = 7.97$, $p < .001$, $\eta_p^2 = 0.22$, with events from Episodes 1 and 10 being rated as the most story-relevant. Events from both Episodes 1 and 10 were numerically more story-relevant than events from any other episode; story-relevance ratings were statistically higher for Episodes 1 and 10 than Episodes 2, 3, 7, and 8 ($ps < 0.006$, surviving Bonferroni correction at $\alpha = 0.006$). Thus participants' judgments of story relevance largely mirrored the beginning and ending effects in memory. Story relevance seems to at least partially explain the serial position effects we observed in the prior experiments.

There was also a main effect of season on judgments of story relevance ($F(7,203) = 16.6$, $p < .001$, $\eta_p^2 = 0.36$). Mirroring what we observed in subjective memory scores, story relevance judgments decreased across the first six seasons, at which point judgments of story relevance increased for Seasons 7 and 8. Events from Season 7 were judged as most story-relevant, with mean relevance judgments significantly higher than all other seasons ($ps < 0.02$, five of which survived Bonferroni correction at $\alpha = 0.007$). This pattern of results also aligns with the pattern of results observed in the earlier over-representation analysis: Seasons 1, 7, and 8 were reliably 'over-represented' on the mental timeline in Experiments 1a and 1b, suggesting a possible link between over-representation and story relevance.

3.3.2.1. Does event memory (Exp.1a) relate to story relevance (Exp.1c)?.

The most direct way to analyze the relationship between event memory (Exp.1a) and storyline (Exp.1c) is to compute the overall correlation between the two. To that end, we ran an across-event correlation relating the mean subjective memory score for each event in Experiment 1a to the mean relevance score for each event in Experiment 1c. We found that these two metrics were reliably correlated ($r = 0.71$, $p < .001$). This pattern is intuitive: Participants are more likely to remember those events that are rated as more story-relevant.

To further probe this question, we next assessed the relationship between the average absolute deviation in placement for each event in Experiment 1a (difference in rank order of the placement and the true event) to the mean story relevance score in Experiment 1c. We observed a reliable negative correlation ($r = -0.36$, $p < .001$), suggesting that more story-relevant events were remembered with relatively greater temporal precision (as indicated by smaller deviations in temporal order memory); note that this pattern is qualitatively identical to what we observed when relating subjective memory score to temporal deviation in Experiment 1a. In contrast, however, relating story relevance to the boundary effects (confusability difference score) observed in Experiment 1a did not reveal a reliable relationship ($r = -0.012$, $p = .90$).

These results highlight a unique puzzle: Are events at the beginnings and endings of seasons better remembered because writers intentionally placed more story-relevant events there? Or are events at the beginnings and endings of seasons judged as more story-relevant because they happened to occur at the beginning or ending of a season? While it may seem obvious that writers would place story-relevant events at beginnings and endings on purpose, there is also strong reason to believe that placement on a timeline per se might influence perceived meaningfulness. Studies on autobiographical memory have shown that people automatically recall events at the beginning and end of the calendar year more often than events in the middle, for instance (Kurbat et al., 1998; Thomsen & Berntsen, 2005). Events that are more likely to be remembered may also become more likely to seem integral in a larger story. In other words: It is possible that arbitrary boundaries, by virtue of influencing what is remembered, also influence perceived meaningfulness. This is something that could be tested directly in future work.

4. Experiment 2: Stranger Things

Unlike *Game of Thrones*, *Stranger Things* has a narrative structure that is much more defined by the individual seasons (i.e., each season has its own narrative with a clear beginning and ending). Would this narrative structure influence event memory? We chose *Stranger Things* because it has a more modular structure, in the sense that each season has a more self-contained narrative. Additionally, because (at the time of the experiment) this story had not yet concluded, it was not yet a fully complete experience. Relative to *Game of Thrones, Stranger Things* consists of relatively few characters and settings.

4.1. Methods

4.1.1. Participants

Per our pre-registered criteria, 9 participants were excluded from this sample; these participants were replaced, yielding a pre-registered sample size of 50. We excluded 2 additional participants who clearly misused the timeline (using the full length of the timeline for each season; See *Supplemental Fig. S3*). Additionally, a sporadic server error led to missing data for some trials (i.e., the data from those trials were never saved). This affected 11 trials total.

4.1.2. Procedure

The procedure was identical to that in Experiment 1a with the following exception. We again tested 2 items per episode but since *Stanger Things* consisted of 34 episodes at the time of data collection, there were 68 real events and 4 fake events (for a total of 72).

4.2. Results and discussion

Again, memory performance was quite high: The average subjective memory score for each real event was 1.52 (SD = 0.29), indicating that for most events participants were somewhere between “Maybe” and “Yes”; the average for the fake events was −0.31 (SD = 0.50).

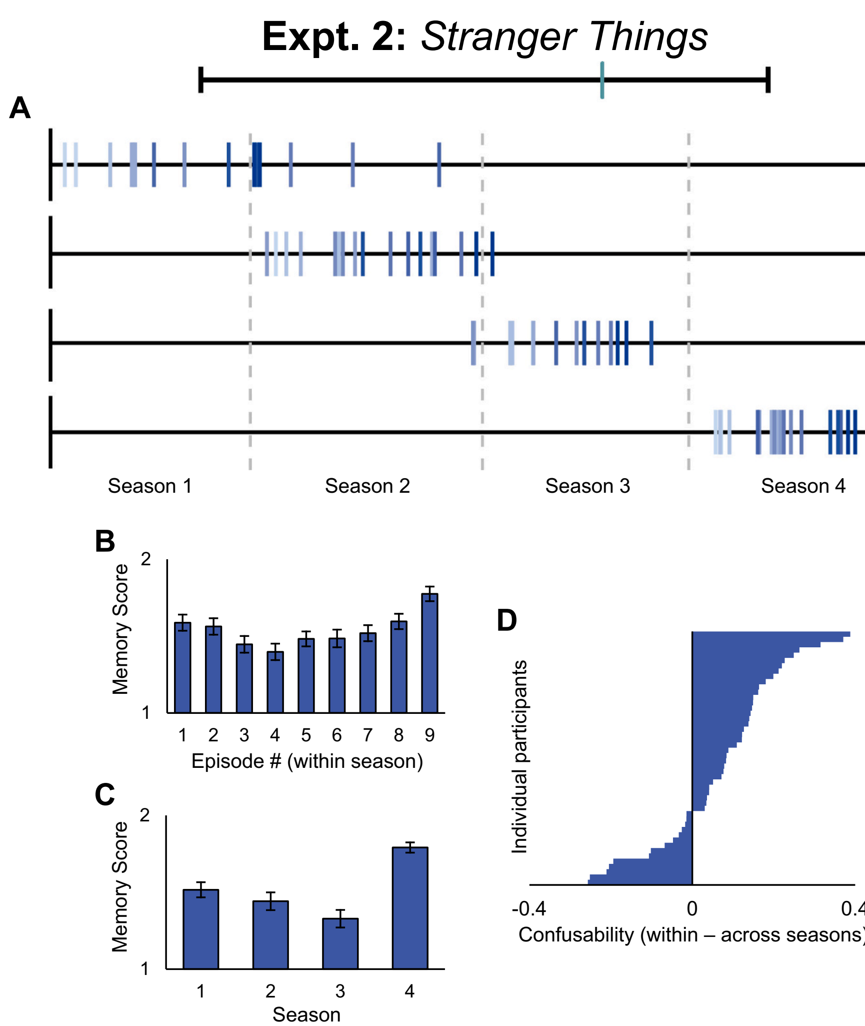

Fig. 6. Results of Experiment 2. (A) Median timeline placement by season (each horizontal line represents a distinct season, from first [top] to last [bottom]) and episode (the earlier episodes in each season are depicted as greyer in color; the later episodes are depicted in a bluer color). (B) The mean subjective memory score for each episode, aggregated across seasons. (C) The mean subjective memory score for each season. (D) The ‘confusability difference score’ (see *Methods*; Fig. 4) for each participant. Error bars represent +/− 1 SE.

Participants were again accurate at placing events on the timeline: The average correlation between participants' reported location and the true serial position of the event was mean $rho = 0.80$ (SD = 0.13; $p < .001$; Supplemental Fig. S3).

4.2.1. Boundaries

Beginning with the effect of structure on memory, a cursory glance at Fig. 6 reveals that seasons were noticeably more separated than the seasons of *Game of Thrones*, although there still appeared to be some overlap between Seasons 1 and 2. Assessing the temporal overlap across seasons, there was only 32% cross-season overlap, a decrease compared to the ~140% overlap for both versions of the *Game of Thrones* study. Most of this overlap came from the first two seasons; the final season was fully separated from the rest. Still, the presence of any overlap suggests that the seasons were not being fully temporally segregated.

We next looked at confusions in the temporal order of events within vs. across seasons. Participants made on average 7.2% (SD = 15.0%) more confusions within vs. across seasons ($t(47) = 3.34$, $p = .002$, $d = 0.48$; Fig. 6D). In other words, as with *Game of Thrones,* boundaries between seasons influenced memory for temporal order.

Relating memory to boundary effects, we observed a reliable, negative correlation between subjective memory score and absolute temporal deviation ($r = -0.25$, $p = .040$). Additionally, there was a reliable, positive correlation between each participant's Spearman rank correlation and their confusability difference score ($r = 0.61$, $p < .001$).

4.2.2. Beginnings/endings

Across all seasons, reported memory steadily decreased from Episode 1 to Episode 4, at which point it steadily increased from Episode 4 to Episode 9. Reflecting this pattern, there was a main effect of episode on memory ($F(8,376) = 9.25$, $p < .001$, $\eta_p^2 = 0.16$; Fig. 6B). Episode 9 was by far the best remembered, significantly higher than all other episodes ($ps < 0.002$, surviving Bonferroni correction at $\alpha = 0.006$). Episode 8, which was the final episode for two of the four seasons, was the second most-accurately-remembered, followed by Episode 1. In other words, there was again clear evidence that beginnings and endings were better remembered than middles.

Across seasons, subjective memory scores steadily decreased each season from the first until the third, at which point there was a large increase in memory for the final season (Main effect of season on memory: $F(3,141) = 36.2$, $p < .001$, $\eta_p^2 = 0.43$; Fig. 6C). Season 4 was the best remembered of all seasons, with reported memory significantly higher than all other seasons ($ps < 0.001$, surviving Bonferroni correction at $\alpha = 0.017$). Events in the first and last seasons were also remembered with the most temporal consistency (lowest standard deviation in placement across participants); fifteen of the 20 most consistently remembered events were from these seasons, greater than chance (permutation test, $p = .006$). Like *Game of Thrones*, there is clear evidence that both beginnings and endings were better remembered — at multiple levels of the event hierarchy, and at both the participant and event level.

We again quantified the amount of 'mental real estate' occupied by each season. The most temporally extended season was Season 1 — but it was not overrepresented. It occupied just about exactly as much space as would be expected ($M = 1.05$, $t(47) = 0.50$, $p = .62$, $d = 0.07$). Seasons 3 and 4, in contrast, were reliably underrepresented ($ps < 0.001$, surviving Bonferroni correction at $\alpha = 0.013$) across participants. One possible explanation is that the most recent seasons of *Stranger Things* aired about 10 months prior to data collection, whereas for *Game of Thrones* the most recent episode had aired about 4 years prior. This raises the possibility that events at the end are initially remembered in a precise, compact way, but that endings take up more 'mental real estate' over time (see also the longitudinal data in Experiment 3).

5. Experiment 3: The Last of Us

The prior experiments offer a clue as to how the structure of event representations changes over time. The events of each season of *Game of Thrones* were temporally extended in participants' responses, whereas the events *Stranger Things* were relatively compressed. Could this be because of a difference in timing? Is it possible that boundaries initially serve to segment one unit from another, but, over time, fade (such that seasons begin to meld together, as happened with *Game of Thrones*)? The show *The Last of Us* presented a unique opportunity to assess event memory longitudinally, allowing us to examine how boundary effects are shaped over time. We collected data from individuals who watched *The Last of Us* within 24 h after the finale of the first season aired (Experiment 3a). We then compared the shape of the same individuals' mental timelines for these events when tested 1 month later (Experiment 3b), and then another 2 months later (Experiment 3c).

5.1. Experiment 3a— Remembering events within 24 h

5.1.1. Methods

5.1.1.1. Participants. 50 new participants completed Experiment 3a (after replacing 5 participants excluded based on the pre-registered exclusion criteria).

5.1.1.2. Procedure. The procedure for Experiment 3a was mostly identical to that of Experiment 1a. However, as *The Last Of Us* consisted of only a single season, we opted to select five events (rather than two events) per episode. As there were 9 total episodes, this meant we had 45 real events and 4 fake events (for a total of 49). *The Last of Us* consists primarily of only two main characters, but the one season we tested covers a range of settings over the course of about one year (excluding a few flashbacks).

5.1.2. Results and discussion

Memory for events within 24 h of the Season 1 finale was quite high: The average subjective memory score for each real event was 1.81 (SD = 0.15), indicating that for most events participants were somewhere between "Maybe" and "Yes"; the average for the fake events was −0.59 (SD = 0.51). The average correlation between participants' reported location and the true serial position of the event was $rho = 0.91$ (SD = 0.076; $p < .001$; Supplemental Fig. S4).

5.1.2.1. Boundaries. Because *The Last of Us* had only a single season at the time of data collection, the primary unit of analysis is no longer seasons, but episodes (we included 5 events per episode instead of 2 to compensate). Fig. 7 reveals that participants placed the events on the timeline almost perfectly. There was only 6.7% cross-episode overlap, a large decrease compared to the over 140% overlap for both versions of the *Game of Thrones* study and even the 38% of *Stranger Things*. The only overlap came from Episodes 4 and 5 (which, of note, contained a single continuous narrative in a single place, unlike most of the other episodes).

We next examined confusability of events within vs. across episodes. For Experiment 3, the 'window' over which this analysis was conducted was ±4 events; we thus excluded the first and last four events from the analysis. We modified the window size for this Experiment to ensure that each event would be compared both to within-unit and across-unit events (given that here we were testing 5, rather than 2, events per episode). Across participants, the average confusability difference score was 17.0% (SD = 11.9%; $t(49) = 10.1$, $p < .001$, $d = 1.43$). As with *Game of Thrones* and *Stranger Things*, boundaries between units (in this case, episodes) influenced memory for temporal order.

Here, unlike with *Game of Thrones* and *Stranger Things*, we did not observe relationships either between subjective memory score and

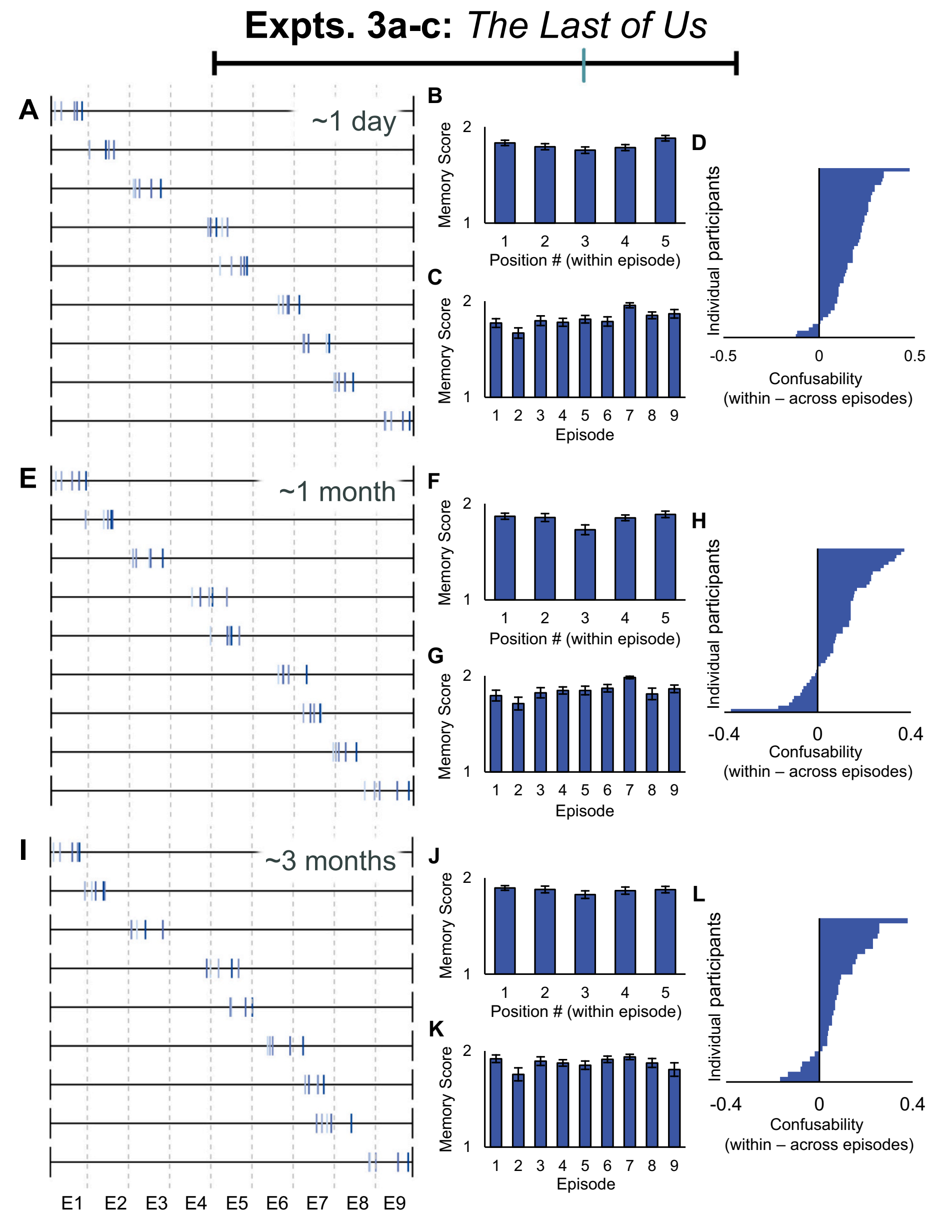

Fig. 7. Results of Experiments 3a (A-D), 3b (*E*-H) and 3c (I-L). (A, E, I) Median timeline placement by episode (each horizontal line represents a distinct episode, from first [top] to last [bottom]) and position (the earlier events in each episode are depicted as greyer in color; the later events are depicted in a bluer color). (B, F, J) The mean subjective memory score for each event position, aggregated across episodes. (C, G, K) The mean subjective memory score for each episode. (D, H, L) The 'confusability difference score' (see *Methods*; Fig. 4) for each participant. Error bars represent +/− 1 SE.

temporal deviation ($r = -0.050$, $p = .74$) or between Spearman rank correlation and confusability ($r = 0.18$, $p = .21$), though we note that this is likely because there was not enough variance in subjective memory score at this timepoint.

5.1.2.2. Beginnings/endings. This is the first data set for which we have sufficient data to look at the order of events within a single episode. Mirroring the effects of event structure that we observed across higher-order units (i.e., seasons), events at the beginning and ending of each episode were better remembered than events in the middle ($F(4,196) = 4.1$, $p = .003$, $\eta_p^2 = 0.08$; Fig. 7B). This provides additional evidence that memories are shaped by event structure at multiple levels of an event hierarchy.

We next turn to the role of different timepoints. Although there was a main effect of episode on memory ($F(8,392) = 4.0$, $p < .001$, $\eta_p^2 = 0.07$), this did not follow the same pattern as the other shows (i.e., memory was not boosted for the beginning and ending of the season; Fig. 7C). Rather, events from Episode 7 had the highest overall subjective memory score (higher than all other episodes, $ps < 0.08$, seven of which survived Bonferroni correction at $\alpha = 0.006$), followed by Episode 9 and then Episode 8. Events from Episode 1 were among the least well-remembered, higher only than Episode 2. Similarly, there were no clear effects of episode on memory consistency (as measured by standard deviation of placement) nor were any of the episodes over-represented on the mental timeline (see *Supplemental Materials*).

5.2. Experiments 3b and 3c — Remembering events 1 and 3 months later

5.2.1. Methods

5.2.1.1. Participants. For Experiments 3b and 3c, we invited back the 50 participants who were included in Experiment 3a. Forty-one of the fifty participated in Experiment 3b, and 32 in Experiment 3c (29 of whom participated also in Experiment 3b). No exclusion criteria were applied to the subsequent samples, given that participants would only be included in this sample if they had passed met the original exclusion criteria. For all analyses comparing Experiments 3a-3c to one another, we only included the participants who were common across the set of experiments of interest.

We additionally collected data from a separate sample of 46 participants tested only at the three-month mark (same time as Experiment 3c; see Experiment S1 in *Supplemental Materials*.) We used the same exclusion criteria as before but did not replace excluded participants (as the goal was to at least match the yield of Experiment 3c).

5.2.1.2. Procedure. The procedure was identical to Experiment 3a. A sporadic server error led to missing data for some trials (4 trials from Experiment 3c and 1 trial from Experiment S1).

5.2.2. Results and discussion

5.2.2.1. Experiment 3b. Participants' memory was remarkably intact one month after the finale (Experiment 3b). In fact, reported memory for the 45 real events increased compared to the 24-h interval in Experiment 3a (from 1.80 to 1.86, $t(40) = 3.2$, $p = .003$, $d = 0.50$). The average correlation between participants' reported location and the true serial position of the event in Experiment 3b was $rho = 0.91$ (SD = 0.79; $p < .001$; Supplemental Fig. S5), which was statistically indistinguishable from participants' correlations at 24 h in Experiment 3a ($p = .86$).

Despite robust memory overall, we observed changes in how the event structure shaped temporal memory between Experiment 3a and 3b. At the group level, the degree of overlap increased from 6.7% to 15.6%. Further, participants' confusability difference scores were reliably lower one month later, decreasing from an average of 17.6% to 11% (SD = 15.4%; $t(40) = 2.9$, $p = .007$, $d = 0.45$; see Fig. 7H). This was driven by an increase in across-episode confusions ($p < .001$), rather than a decrease in within-episode confusions ($p = .32$). Additionally, we observed some evidence for a relationship between memory and boundary effects in Experiment 3b, in contrast to what we observed at 1 day in Experiment 3a (see *Supplemental Materials*). In other words, we observed most of what we would expect to find if the boundaries 'faded' over time. These findings cannot be explained simply by memory decay, as subjective memory scores were actually *higher* after 1 month than after 24 h.

Although we observed changes in how event structure shaped memory, there were comparable patterns for beginnings and endings (Fig. 7F-G; see *Supplemental Materials*). Further, there was no clear pattern with respect to the 'mental real estate' analysis. As observed at 24 h, no episode was over-represented on the timeline (see *Supplemental Materials*).

5.2.2.2. Experiment 3c. Three months after the finale (Experiment 3c), memory remained remarkably intact (mean $rho = 0.89$; SD = 0.16; $p < .001$; see also *Supplemental Materials* and Supplemental Fig. S6) and remained indistinguishable from the correlations at 24 h and 1 month (pairwise $ps > 0.12$). This may not be surprising insofar as some prior work has shown remarkable retention for narrative details over durations much longer than three months (Doolen & Radvansky, 2022; Furman, Dorfman, Hasson, Davachi, & Dudai, 2007; Wagenaar, 1986). However, we also observed the same patterns of memory degradation that we observed after one month: The degree of overlap increased from 15.6% to 22.2%, confusability difference scores continued to decrease, to 8.6% (SD = 12.1%) on average (this was a numerical decrease relative to 1-month: $t(28) = 1.2$, $p = .26$, $d = 0.21$, and a significant decrease relative to 24 h: $t(31) = 4.1$, $p < .001$, $d = 0.72$; see Fig. 7L). These results suggest that memories for complex events degrade in a systematic, predictable way, with respect to their event structure. Furthermore, averaging across episodes, each episode tended to occupy more mental real estate each subsequent time memory was tested (though we note that the pairwise differences do not reach significance; Exp 3a $M = 0.44$; Exp 3b $M = 0.53$; Exp 3c $M = 0.69$; Exp 3a vs. Exp 3b: $p = .50$; Exp 3b vs. Exp 3c: $p = .27$; Exp 3a vs. Exp 3c: $p = .10$).

5.2.2.3. Effects of repeated testing?. To ensure that nothing about the results of Experiment 3c was specific to participants who had repeatedly participated in the study, we collected data from a separate sample of participants tested only at the three-month mark (same time as Experiment 3c; see Experiment S1 in *Supplemental Materials*.) All meaningful patterns from Experiment 3c replicated (see *Supplemental Materials*; Supplemental Figs. S7, S8A). Note that that subjective memory scores in this independent sample were slightly lower than in the longitudinal sample, but still about as high as in the original sample collected after one day (Experiment 3a). This suggests that the improved subjective memory scores in Experiments 3b and 3c may be due to a sort of 'reactivation' effect caused by repeated testing of the same items. Crucially, however, this does not affect the main patterns of interest in our study. For a more complete comparison of this independent sample and the three longitudinal samples, see Supplemental Fig. S8B–D.

6. General discussion

Here, we explored event representations at the scale of ordinary experience by probing people's mental timelines for three popular television shows. Three key patterns were consistent across the shows that we tested: (1) Events *within* units (seasons, episodes) were more readily confused than events *across* units; (2) Individual units (e.g., seasons) were expanded, rather than contracted, on the mental timeline, and (3) Events at the beginnings and endings of multiple levels of the event hierarchy were better remembered. These patterns are hallmarks of event-related cognition. Thus, these findings provide evidence that, even on longer timescales, and for more complex events, memories are organized with respect to their event structure.

There are many patterns in these data that are not straightforwardly predicted by prior event-related work on shorter timescales. We see *some* markers of boundary effects (i.e., confusability within vs. across units), but not others. For instance, we may have expected that events from different units (seasons, episodes) would be temporally segregated to a greater extent or that there would be greater confusability across vs. within units (Clewett et al., 2020; DuBrow & Davachi, 2013; Ezzyat & Davachi, 2014). In fact, we largely found the opposite: Seasons appeared to overlap with one another to a large extent and we observed greater confusability within units. For *The Last of Us*, we did observe the sort of temporal segmentation between episodes that we may have expected when participants were tested 24 h after the fact. Yet when we tested participants 1 month and then 3 months later, we saw that this segmentation slowly faded. Collectively, these data provide initial evidence that boundary effects, like temporal segmentation, fade over time. This conclusion is not obvious. It could well have been that segmentation increased over time, such that distinct units became increasingly temporally segmented.

That said, even beyond the difference in timescale, there are many differences between our approach and the approach(es) of prior work that limit generalization across the two. First, the gaps between episodes and seasons are (usually) filled with other life experiences, whereas the boundaries in prior work (Clewett et al., 2020; DuBrow & Davachi, 2013; Ezzyat & Davachi, 2014) contain no additional information. Second, television shows may be called to mind in a variety of different ways across time, either via explicit recaps built into the show, conversations with friends, or spontaneous recollection. These 'recursive remindings' may improve overall memory (Jacoby & Wahlheim, 2013), and perhaps the way that memories are organized with respect to their event structure. Third, the events of a television series are obviously more complex than simple images on a computer screen, requiring significantly more memory resources. This also means that models of event segmentation that focus on working memory are unlikely to explain the results observed here. Finally, as we mentioned in discussing the results of Experiment 1a, the events here are intertwined as part of a broader narrative structure (rather than being arbitrarily ordered, unrelated items). Through a different lens, however, these caveats are also strengths of the current approach: We have demonstrated that event representations can be studied at the scale of ordinary experience despite the messiness and complexity of the real world.

6.1. Does Game of Thrones reflect ordinary human experience?

One might argue that these results reveal less about event representation for ordinary experiences and more about the narrative structure of a unique stimulus (but note that much prior work on event representation has relied on similar stimuli, although on a shorter timescale; see Baldassano et al., 2017; Copeland et al., 2009; Doolen & Radvansky, 2021, 2022; Speer et al., 2007; Zacks et al., 2010). After all, television shows not only have defined boundaries, but story arcs that — cliffhangers aside — typically map onto these boundaries in a structured way (i.e., new narratives that begin and end each season). As we foreshadowed in the introduction, though, it seems to us that narrative elements in life also map onto event boundaries in meaningful ways. A huge job promotion might occur just before a move across the country (because the promotion might *cause* a move across the country); a new life in a new home might therefore coincide with the start of a new relationship. And in life, as in television shows, some narratives begin, are interrupted, and begin again. Some characters may disappear only to make a dramatic, unexpected return. Narrative elements are complex and interwoven such that sometimes, meaningful story arcs are only realized after retrospection (see Conway, 1996 for a discussion of how autobiographical memory is shaped by these factors). Television shows obviously cannot capture the true nature of human experience (but see Yang et al., 2022), but they are a useful place to get started.

So in what way do the effects observed here reflect event representation? It is possible that we observe boundary effects because of narrative structure that happens to be confounded with season and episode boundaries. It is also possible that we observe boundary effects because of the pauses between seasons and episodes (i.e., the simple fact that most viewers watch one episode every week or one season every year, regardless of how the narrative unfolds). There is an ambiguity here that is difficult to resolve.

We have several thoughts about the ambiguity between narrative structure and temporal structure. First, we think it is important to note that events in life are similarly structured: Semesters are often the main unit of an academic calendar, but there are also breaks between them. Football seasons are broken up by practices and bye-weeks. Our work lives and our home lives are intermixed. If our goal is to truly understand "event representation at the scale of ordinary experience", we must accept, to some degree, that "narrative structure" and "pauses" are often confounded. Second, we would note that even carefully controlled lab experiments often have a similar ambiguity. After decades of study, it remains unresolved whether event boundary effects are fundamentally about context, or instead about prediction error, or inferred causal structure, or some sort of event representations that exist independently of each of these things (see, e.g., Yates et al., 2023). Third, we collected data examining to what extent subjective story-relevance of events may explain other effects that we observe (Experiment 1c). Perhaps unsurprisingly, story-relevance does help to explain some effects. For instance, there is a robust relation between memory and story-relevance. This could explain why events at the beginnings and endings of events are better remembered — because story writers purposefully place more critical events at the beginnings and endings of seasons. Yet the opposite could be true as well. It could be that events are better remembered, and thus perceived as more story-relevant, *because* they happen to occur at the beginning or ending of a season. This is true in life, as well. If one gets off the plane in a new city with the aim of starting a new life, the first restaurant they visit may have a special importance — if only because it happened to occur in that new life stage. A chance encounter with a stranger around that time may seem like something more, if only because of when it happened in a larger story. Indeed, prior work has shown that people are more likely to freely recall events from the beginning and ending of the calendar year (Kurbat et al., 1998; Thomsen & Berntsen, 2005) and that autobiographical memory is similarly influenced by certain "landmark" events (see Shum, 1998).

One may also wonder about the inconsistencies across the shows that we tested. We see the diversity in the shows we tested as a strength: The fact that some patterns were consistently discernible across three highly unique narratives (some experienced over the course of a couple months, and some over nearly a decade) speaks to the generality of the findings here.

There are a few specific questions we hope future work will address. *First*: To what extent are mental timelines influenced by how the information was encoded in the first place? Does it matter, for instance, if someone watched an entire show over the course of a month versus over the course of ten years? *Second*: How are mental timelines influenced by encoding the same information a second, or third, or fourth time? Do people's timelines become more precise? Are boundary effects reinforced each time the information is re-encoded? How do trailers and recaps influence temporal organization? *Third*: To what extent do these results reflect narrative structure per se, and to what extent would they generalize to more natural events (without such clear beginnings and endings, for instance)? This list of questions is not exhaustive. We strongly suspect that there are many opportunities here for further study.

7. Conclusion

We remember not only *what* happens in our lives, but *when*. Our memories are embedded within a complex structure of events — complete with interwoven narratives, triumphs and failures, cliffhangers, and plot twists. Yet while the study of event cognition has grown in prominence over the last two decades, much of what we know about how we represent the structure of experience is confined to the timescale of seconds, minutes, or hours at most. Here, we took a foray into the study of hierarchically structured events on the scale of months and years. Taking memories for popular television shows as a case study, we showed that many hallmarks of event cognition — namely, boundary effects and the prioritization of beginnings and endings — manifest on larger and more complex scales. These new data open the door to an array of new questions and offer the promise of understanding event cognition at the scale of ordinary human lives.

CRediT authorship contribution statement

Sami R. Yousif: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sarah Hye-yeon Lee:** Writing – review & editing, Methodology, Conceptualization. **Brynn E. Sherman:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Formal analysis. **Anna Papafragou:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

None.

Data availability

All of our data are available on an OSF page: https://osf.io/syw84/

Acknowledgements

For valuable feedback on this project, the authors thank Tristan Yates, the members of Josh Knobe's lab group at Yale University, and the members of the Language & Cognition Lab at the University of Pennsylvania. This work was supported by the National Science Foundation (Grant #2041171, to Anna Papafragou; Grant #SMA-2313948, to Brynn Sherman), and the MindCORE Program at the University of Pennsylvania (via an award to Sami Yousif).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.105833.

References

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron, 95*(3), 709–721.

Bangert, A. S., Kurby, C. A., Hughes, A. S., & Carrasco, O. (2020). Crossing event boundaries changes prospective perceptions of temporal length and proximity. *Attention, Perception, & Psychophysics, 82*, 1459–1472.

Bruner, J. S. (1994). The "remembered" self. In U. Neisser, & R. Fivush (Eds.), *The remembering self* (pp. 41–54). New York: Cambridge University Press.

Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications, 11*(1), 1–14.

Conway, M. A. (1996). Autobiographical memory. In *Memory* (pp. 165–194). Academic Press.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*, 261–288.

Conway, M. A., & Rubin, D. C. (2019). The structure of autobiographical memory. In *Theories of memory* (pp. 103–137). Psychology Press.

Copeland, D. E., Radvansky, G. A., & Goodwin, K. A. (2009). A novel study: Forgetting curves and the reminiscence bump. *Memory, 17*, 323–336.

Doolen, A. C., & Radvansky, G. A. (2021). A novel study: Long-lasting event memory. *Memory, 29*, 963–982.

Doolen, A. C., & Radvansky, G. A. (2022). A novel study: Hypermnesia for books read years ago. *Memory, 30*, 92–103.

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General, 142*(4), 1277–1286.

DuBrow, S., & Davachi, L. (2014). Temporal memory is shaped by encoding stability and intervening item reactivation. *Journal of Neuroscience, 34*, 13998–14005.

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory, 134*, 107–114.

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science, 22*, 243–252.

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron, 81*, 1179–1189.

Ezzyat, Y., & Davachi, L. (2021). Neural evidence for representational persistence within events. *Journal of Neuroscience, 41*, 7909–7920.

Faber, M., & Gennari, S. P. (2017). Effects of learned episodic event structure on prospective duration judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1203–1214.

Furman, O., Dorfman, N., Hasson, U., Davachi, L., & Dudai, Y. (2007). They saw a movie: Long-term memory for an extended audiovisual narrative. *Learning & Memory, 14*, 457–467.

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 1075–1090.

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition, 41*, 625–637.

Ji, Y., & Papafragou, A. (2020a). Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition, 197*, Article 104197.

Ji, Y., & Papafragou, A. (2020b). Midpoints, endpoints and the cognitive structure of events. *Language, Cognition and Neuroscience, 35*, 1465–1479.

Ji, Y., & Papafragou, A. (2022). Boundedness in event cognition: Viewers spontaneously represent the temporal texture of events. *Journal of Memory and Language, 127*, Article 104353.

Kahana, M. J., Diamond, N. B., & Aka, A. (2022). Laws of human memory. *PsyArXiv*. https://doi.org/10.31234/osf.io/aczu9

Kurbat, M. A., Shevell, S. K., & Rips, L. J. (1998). A year's memories: The calendar effect in autobiographical recall. *Memory & Cognition, 26*, 532–552.

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition, 40*, 812–826.

Liverence, B. M., & Scholl, B. J. (2012). Discrete events as units of perceived time. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 549–554.

McCrary, J. W., Jr., & Hunter, W. S. (1953). Serial position curves in verbal learning. *Science, 117*(3032), 131–134.

McNerney, M. W., Goodwin, K. A., & Radvansky, G. A. (2011). A novel study: A situation model analysis of reading times. *Discourse Processes, 48*, 453–474.

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology: General, 64*, 482–488.

Murphy, G. (2004). *The big book of concepts*. MIT press.

Pradhan, R., & Kumar, D. (2021). Event segmentation and event boundary advantage: Role of attention and postencoding processing. *Journal of Experimental Psychology: General, 151*, 1542–1555.

Pu, Y., Kong, X. Z., Ranganath, C., & Melloni, L. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature Communications, 13* (1), 1–13.

Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science, 21*, 269–272.

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition, 34*, 1150–1156.

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *Quarterly Journal of Experimental Psychology, 64*, 1632–1645.

Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychonomic Bulletin & Review, 17*, 900–904.

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 608–620.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences, 17*, 133–140.

Roediger, H. L., III, & Crowder, R. G. (1976). A serial position effect in recall of United States presidents. *Bulletin of the Psychonomic Society, 8*, 275–278.

Rouhani, N., Stanley, D., COVID-Dynamic Team, & Adolphs, R. (2023). Collective events and individual affect shape autobiographical memory. *Proceedings of the National Academy of Sciences, 120*(29), Article e2221919120.

Sehulster, J. R. (1989). Content and temporal structure of autobiographical knowledge: Remembering twenty-five seasons at the metropolitan opera. *Memory & Cognition, 17*, 590–606.

Sherman, B. E., DuBrow, S., Winawer, J., & Davachi, L. (2023). Mnemonic content and hippocampal patterns shape judgments of time. *Psychological Science, 34*, 221–237.

Shipley, T. F., & Zacks, J. M. (Eds.). (2008). *Understanding events: From perception to action*. Oxford University Press.

Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin, 124*, 423–442.

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science, 18*, 449–455.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138*, 236–257.

Teigen, K. H., Böhm, G., Bruckmüller, S., Hegarty, P., & Luminet, O. (2017). Long live the king! Beginnings loom larger than endings of past and recurrent events. *Cognition, 163*, 26–41.

Thomsen, D. K., & Berntsen, D. (2005). The end point effect in autobiographical memory: More than a calendar is needed. *Memory, 13*, 846–861.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18*, 225–252.

Wen, T., & Egner, T. (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *Cognition, 225*, Article 105145.

Yang, B. W., Deffler, S. A., & Marsh, E. J. (2022). A comparison of memories of fiction and autobiographical memories. *Journal of Experimental Psychology: General, 151*, 1089–1106.

Yates, T. S., Sherman, B. E., & Yousif, S. R. (2023). More than a 'moment': What does it mean to call something an event? *Psychonomic Bulletin & Review, 30*, 2067–2082.

Yates, T. S., Skalaban, L. J., Ellis, C. T., Bracher, A. J., Baldassano, C., & Turk-Browne, N. B. (2022). Neural event segmentation of continuous experience in human infants. *Proceedings of the National Academy of Sciences, 119*(43), Article e2200257119.

Yousif, S. R., & Scholl, B. J. (2019). The one-is-more illusion: Sets of discrete objects appear less extended than equivalent continuous entities in both space and time. *Cognition, 185*, 121–130.

Zacks, J. M. (2020). Event perception and memory. *Annual Review of Psychology, 71*, 165–191.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General, 138*, 307–327.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*, 273–293.

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience, 4*, 1–15.