Retrospective duration judgments of naturalistic events depend on memories of event boundaries

Winny W. Y. Yue[1], Jing Liu[2], Ziqing Yao[1], Yuqi Zhang[1], Zexuan Mu[1],
and Xiaoqing Hu[1,3*]

1, Department of Psychology, The State Key Laboratory of Brain and Cognitive Sciences,
The University of Hong Kong, Hong Kong SAR, China
2, Department of Applied Social Sciences,
The Hong Kong Polytechnic University, Hong Kong SAR, China
3, HKU-Shenzhen Institute of Research and Innovation, Shenzhen, China

* Correspondence should be sent to:
Xiaoqing Hu
Department of Psychology, The University of Hong Kong,
Pokfulam, Hong Kong SAR, China
Email: xiaoqinghu@hku.hk

Retrospective duration judgments of naturalistic events depend on memories of event boundaries

Abstract

Daily planning and goal-directed behavior rely on accurate judgments of the duration of past experience. Although retrospective duration judgments are often shorter than the actual time elapsed, how episodic memory changes may impact duration judgments remains unclear. Here, participants watched videos depicting daily events with clear boundaries segmenting each subevent. Participants then completed recall and duration judgment tasks both immediately and after 7 days. Results showed that the recall of the event structure, specifically the number of subevents, significantly influenced immediate and delayed duration judgments. In contrast, memories of gist and number of details had no major impact. However, subevent duration judgments differ, with immediate judgments linked to gist and detail richness, while delayed judgments tend to average out. Together, these results provide new knowledge on the relationship between retrospective duration judgments and memories of naturalistic events, and how such relationship changes over time for different event structures.

Keywords:

Retrospective duration judgements, naturalistic events, event boundaries, memory hierarchy

Introduction

People can perceive event durations spanning from mere milliseconds to several years (Karmarkar & Buonomano, 2007). Judgment of an event duration can be conducted retrospectively, either immediately following an experience or after a considerable lapse of time (Grondin et al., 2014; MacDonald, 2014; Ornstein, 1969; Tsao et al., 2022; Yarmey & Matthys, 1990; Zakay & Block, 2004; Zakay & Fallach, 1984). Our ability to assess temporal duration of past experience is crucial for various cognitive processes, including motor learning, strategic planning, as well as speech production and interpretation (Gransier et al., 2023; Little et al., 2013; Paton & Buonomano, 2018; Zion Golumbic et al., 2012). Despite its importance, our retrospective duration judgment is not always accurate; it tends to be shorter compared to the actual duration of the event, known as temporal compression (MacDonald, 2014; Tsao et al., 2022; Zakay & Block, 2004). Concurrent with this temporal compression, our memories of past experience also undergo significant changes, such as episodic forgetting and memory schematization (Radvansky et al., 2022; Santoro et al., 2016). However, whether and how retrospective duration judgments are derived from our recall of specific episodic memory content remains elusive.

An interesting observation about retrospective duration judgment is that richer memories were often associated with less compressed and more accurate assessment of the event's duration (Block, 1992; Clewett et al., 2019; D'Argembeau et al., 2022; Jeunehomme et al., 2018; Jeunehomme & D'Argembeau, 2019; Zakay & Block, 2004). These results support the storage-size hypothesis, which posits that people judge duration based on the volume of retrievable memory content (Ornstein, 1969). Notably, for naturalistic event, memory is better encoded at event boundaries, making these boundaries important anchor points during memory retrieval (Michelmann et al., 2023). Indeed, boundaries typically involve contextual changes like shifts in emotional state and environmental transitions (e.g., moving from one room to another), and they are marked by greater neural pattern differences indicating the update and reset of memory representations at these critical points (Bangert et al., 2020; Block, 1992; Ezzyat & Davachi, 2014; Lositsky et al., 2016; Swallow et al., 2009; Zakay & Block, 2004). In contrast to memories across boundaries, memories of events occurring between boundaries is less reliable and is believed to be integrated into a single coherent unit before being consolidated for long-term storage (Ben-Yakov et al., 2013; Dudai et al., 2015; Sun et al., 2020; Terada et al., 2017; Wallenstein et al., 1998). This integration might also influence our retrospective duration judgment in that items spanning across event boundaries are perceived as temporally more distant compared to items within the same boundary (Pu et al., 2022).

Does this phenomenon imply that when we assess event duration, we would primarily rely on event boundaries, neglecting the granular information abounding within event segments? To address this question, we used a well-defined memory structure that includes not only sharp event boundaries that segment individual subevents, but also distinct contextual information. We then investigate whether durational judgement would be correlated with memory changes at different hierarchical levels, such as the number of boundaries, gist versus details for events and for each subevents.

In fact, there has been a previous attempt to explore how memory structure might influence duration judgment, although limited to only one study. By examining the relationship between participants' memories of naturalistic events (a campus walk) and their duration judgments, this study suggests that the number of event segments predicted the duration judgments while the details within those segments did not (Jeunehomme & D'Argembeau, 2019). However, one important caveat of this study is that event boundaries are not well-defined and are not reported in real time during encoding and retrieval. Instead, participants were required to engage in a detailed mental replay of the event prior to verbal recall, and boundaries were identified by external coders using transitional words in the verbal recall (e.g., "then," "next"). Because detailed mental replay encouraged vivid and continuous recall, event boundaries identified during retreival are likely different from those detected during encoding. Some other studies examined the changes in reported duration over a period of delay, but these work did not examine the significance of hierarchical structure of memory such as events and their subevents segmented by boundaries (Grondin et al., 2014; Yarmey & Matthys, 1990). Most of these studies also adopted a between-subject design, leading to high variability in reported durations (Lositsky et al., 2016; Safi et al., 2024).

We wanted to induce memory changes within subjects so as to look at the possible change in durational judgements. Recognizing that forgetting naturally occurs over time, we designed a within-subject test-retest study with a 7-day test interval. To examine the memory changes structurally, we edited our video stimuli to incorporate sharp event boundaries that separate subevents with distinct gist and details. We anticipate that duration judgments would be primarily driven by memories of event boundaries, while more granular memories such as gists and details will have smaller contributions to duration judgment.

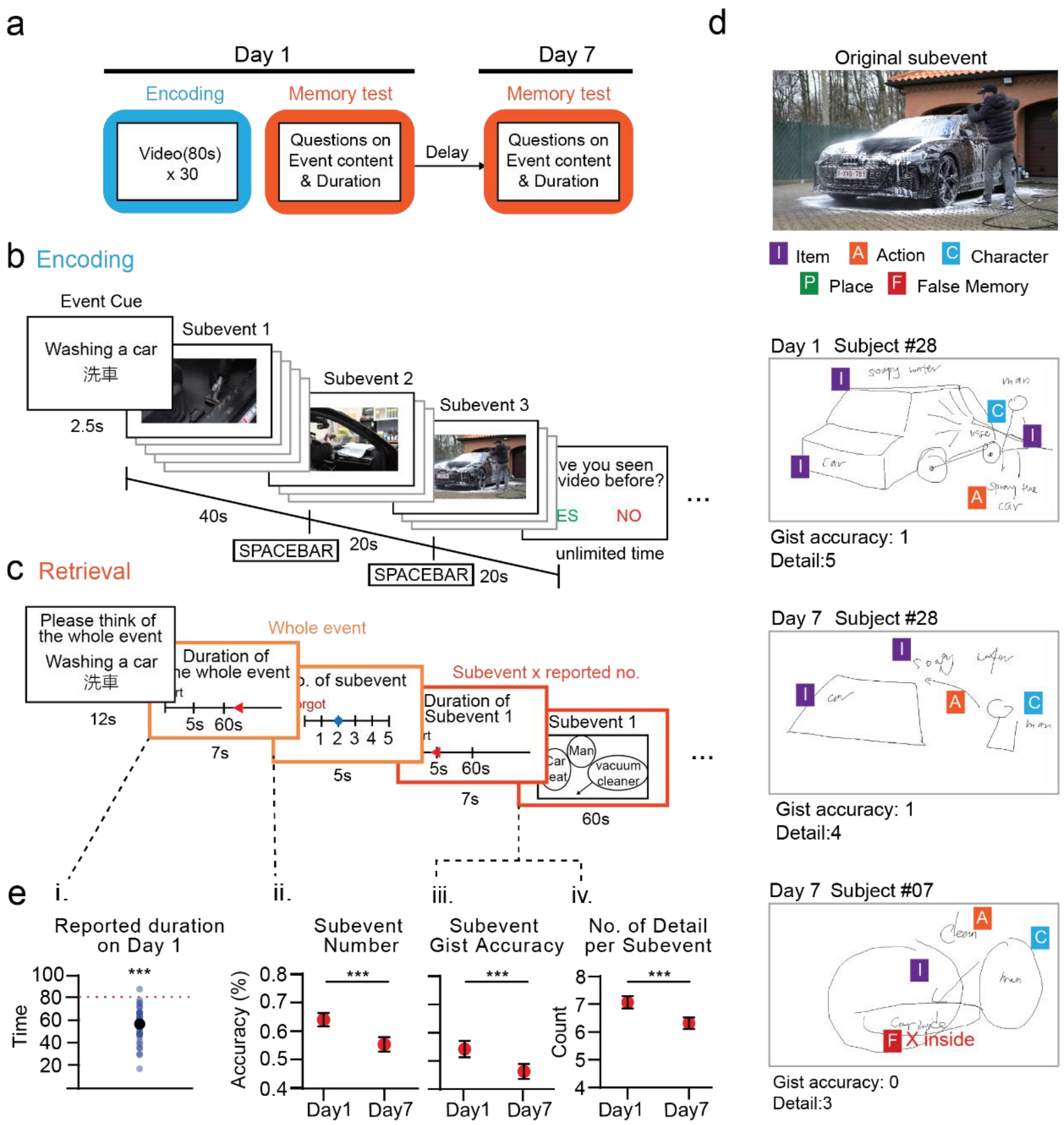

Fig. 1. Experimental paradigm. (a) Experimental flow. (b) The encoding phase. Participants viewed thirty 80-second videos depicting daily life events, each comprising 2-4 nested subevents each lasting either 20 or 40 seconds. A short phrase describing the event was presented before each video as a cue. During video watching, participants were instructed to press the spacebar to indicate event boundaries when context or activity shifts. (c) The retrieval phase. Prompted by cues, participants were asked to recall the corresponding video, followed by questions about the entire event and each subevent. Specifically, participants reported the duration of the whole event and of each subevent, reported the numbers of subevent in each individual event, and drew each subevent out with labels to show the remembered contents. (d) Example coding for drawings by participants. Coders first assessed whether the drawings captured the gist of the subevent, then categorized and counted the number of details the

participants included. (e)(i) Participants' duration judgments on Day 1 showed temporal compression. Blue dots represent the mean reported duration for each participant. Black dots indicate the mean reported duration for all participants. Red dotted line indicates actual event duration. (ii-iv) Changes in memory attributes across days (ii) Reported subevent number (iii) Recalled gist accuracy (iv) Number of details. Participants' recall of the events significantly declined from Day 1 to Day 7, i.e., forgetting. ($n = 48$). The dataset excluded invalid responses and empty responses (see Methods). See Fig. S2 for plots with memory performance before filtering. Error bars show the SEM.

*** $p < .001$ (all two-tailed).

Method

Data and relevant code are available on OSF at https://osf.io/m43yh/?view_only=8f93a14ba974440f91c3be8a4e9c99ad

Participants

Fifty-five individuals (Female = 39, Preferred not to say = 1; $M_{age} = 22.589$; $SD_{age} = 3.109$) were recruited from the University of Hong Kong for this experiment. The sample size was determined with reference to prior between-subject studies on retrospective duration and episodic memories (Furman et al., 2007; Jeunehomme et al., 2018). Anticipating potential challenges such as poor memory performance and participant absenteeism in the delayed experiment, we recruited more participants. Ultimately, six participants were excluded due to absenteeism in the delayed experiment, and one participant was excluded for low performance during the filtering process (see Analysis section). This resulted in a final sample size of 48 participants in the analyses reported here.

All eligible participants had normal or corrected-to-normal vision and hearing, were not color-blind, and had no chronic medical conditions, history of severe mental illness, neurological disorders, or current clinical diagnosis for any psychiatric conditions. The experiments were conducted in either simplified or traditional Chinese, with all instructions and experimental materials remaining consistent across both language versions. Participants were assigned to the language condition based on their preferred language. All participants provided written informed consent and received monetary compensation upon completing the entire experiment ($250 HKD, approximately $32 USD). This study has been approved by and conforms to the standards of the Human Research Ethics Committee for Non-Clinical Faculties at the University of Hong Kong (Ethics Approval No.: EA210341).

Procedures

To investigate the role of memory in event duration judgment, we conducted a within-subject experiment over 7 days, tracking how alterations in memory at different structural levels influence participants' duration assessments. Memory tests were administered immediately after participants watched the videos on Day 1 and again on Day 7 (see Fig. 1a). During these tests, participants described their memories in the form of drawings and reported duration of each event and subevent (See Figs. 1b, c). We then systematically coded their memories and analyzed both high-level structure (i.e., the number of subevents) and low-level features, including gist as well as the quantity and categories of details for each recalled subevent (see Fig. 1d).

Given that our study primarily focuses on assessing retrospective duration judgments, participants were instructed to remove all wearable watches and turn off their electronic devices before the experiment began. To eliminate any potential time cues, all clocks in the lab and on the computer interface were removed. Any questions regarding the time were not answered to maintain the integrity of duration judgements. Participants were informed at the time of sign-up that the experiment would conclude on schedule, lasting approximately 2.5 hours for Day 1 and 1.5 hours for Day 7. Additionally, they were not informed of the number of tasks or the duration of each task before and during the experiment.

Encoding session

During the encoding session, participants watched 30 videos depicting distinct daily life events. Before the task began, they were briefed on the structure of the events, with subevents defined as distinct segments representing different activities. Participants were informed that each event would consist of a varying number of subevents. They were instructed to engage with the videos attentively, as if personally experiencing the events. However, since the study focused on naturalistic recall, participants were neither directed to memorize the videos nor informed of an upcoming memory retrieval task.

To assess participants' ability to recognize the transition between subevents and thereby confirm their understanding of the event structure, they were instructed to press the SPACEBAR as soon as they recognized a change in subevents while watching the videos. After each video, participants were asked whether they had seen the video before. Trials with an unmatched number of presses or where participants had previously viewed the video were excluded from the analysis to ensure data integrity.

Before each video, a cue word was presented for 2.5 seconds to serve as a prompt for recall during the subsequent retrieval task. This cue word encapsulated the overall event, such as "washing car," without specifying any subevents.

The task lasted approximately 35 to 45 minutes, with several untimed breaks included. Before the main task, participants completed a practice session that included a demonstration and three practice trials.

Day 1 Retrieval Test

Immediately following the encoding session, participants underwent a surprise test designed to structurally assess their baseline memory and duration judgments for each event and subevent. All 30 videos from the encoding session were included in this test. Each trial consisted of three parts: first, participants estimated the duration of the entire event. Next, they identified the number of subevents within the video. Finally, they provided the duration and described the content of each subevent.

In the first part of the test, participants were asked to report the total duration of the entire video using a slider. To prevent participants from recalling exact durations based on specific numbers or slider positions, their responses were shown without any numerical indicators. Additionally, reference points were provided only at '5s' and '60s'on the slider, which ranged from 0 to 300 seconds, to assist with estimation.

Previous studies permitted participants unlimited time to mentally replay the event in detail, potentially resulting in duration estimates that reflect a prospective perception of their mental replay rather than a retrospective judgment. In contrast, our study imposed a fixed 12-second time limit for participants to form their judgment, followed by an additional 7 seconds to report their answer. During this phase, participants were instructed to focus solely on the current question and not to consider other videos or forthcoming questions.

Next, to evaluate higher-order memory of the event structure, participants reported the number of subevents they recalled from the video. Participants were instructed to honestly indicate the number of subevents they could remember, even if they had forgotten some content or details, using a slider ranging from 1 to 5.

To assess whether lower-order memory contributed to duration judgments, participants were asked to report both the duration and content of each subevent. They provided the duration using the same slider as for the total event duration. Following this, participants were instructed to draw and label the scene of each subevent. Drawing was selected over verbal descriptions to minimize cognitive effort, given the study's focus on visual stimuli. Participants depicted objects, characters, and background details on a blank page, based on their recollections, and used arrows and labels to indicate characters' actions. Participants were informed that they could use clear labeling with circles instead of creating detailed drawings.

Participants had one minute to complete each subevent before moving on to the next. To ensure understanding and accuracy, they practiced using the drawing pad under the experimenter's guidance during a demonstration and practice session prior to the actual task.

Throughout the task, participants who were uncertain of their responses or had minimal recall were permitted to skip the specific questions they were unsure about. Trials were fully randomized, and an untimed break was provided after every five videos. The entire task lasted approximately 1 to 1.5 hours. At the end of the experiment, participants were free to leave.

Day 7 Retrieval Test

To analyze changes in memory and reported durations, participants repeated the memory test after a 7-day delay, anticipating forgetting. This 7-day interval was determined through a pilot test to ensure sufficient trials with natural forgetting. Trials were fully randomized, and all videos from the initial test were included in this follow-up assessment.

Materials

In our study, we utilized 34 short videos, each uniformly lasting 80 seconds. These videos depict various everyday activities, for example washing a car and playing at the beach. To maintain a clear event structure within each video, we edited them to generate distinct event boundaries with noticeable changes in activity and/or environment. Each video was divided into 2 to 4 subevents, which correspond to natural and meaningful segments of a complex event, as defined by Swallow, Zacks, and Abrams (2009). For example, in the video depicting 'washing a car' the subevents included 'vacuuming the seats', 'cleaning the window' and 'hosing off the car' (see Fig. 1b). We varied the subevents' durations between 20 and 40 seconds to introduce variation, thereby preventing participants from perceiving all subevents as having the same length. This approach aimed to prevent participants from calculating the total event duration based on a fixed subevent length during the retrieval test, while also ensuring participants had sufficient time to fully perceive the activities.

These videos, sourced from platforms like YouTube and Bilibili (a complete video list is provided in the supplementary materials). To maintain consistency and minimize external distractions, each subevent within an event was filmed in a single continuous shot, with no manual editing, visual effects, or subtitles. Additionally, any logos in the videos were blurred to avoid biasing participants' perceptions. All videos were edited in Adobe After Effect and were exported with Adobe Media Encoder to resolution of 1920 (w) x 1080 (h).

These videos were tested in a pilot study involving three coders who assessed event boundary consistency. This preliminary testing was conducted to minimize variations in the perception of event boundaries during the actual experiment. Videos with unclear event boundaries or ambiguous content were excluded from the final experiment. To familiarize participants with the procedure, four of these videos were used in a practice session. The remaining 30 videos were presented during the encoding session of the task. During the study, we replaced two videos, numbered 9 and 18, due to ambiguous event boundary identification observed in the initial experimental phase. Consequently, the trials associated with these videos from the first eight participants were excluded from the final analysis to ensure the reliability and consistency of the data.

The experiment was conducted using the software PsychoPy version 2022.2.4. During the retrieval test, participants were instructed to use a drawing pad (Wacom Intuos CTH-680) to complete the drawing and labeling tasks.

Data analysis

Exclusion criteria

One participant was excluded due to low performance on the retrieval test. To ensure the accuracy of the data, specific criteria were applied to filter out certain trials. These criteria included: (1) trials lacking subevent indications or with incorrect indications during video watching, to ensure a clear understanding of the event structure, and (2) trials where participants reported having seen the video before, to prevent prior knowledge from biasing the results. To accurately align event content memories with duration judgments, the following trials were excluded: (1) trials with skips during the subevent number recall period, (2) trials with no response or a skip during the total duration report, and (3) trials where participants responded with 'I forgot' or did not provide an answer during the subevent number report. Additionally, to ensure the validity of responses, the analysis excluded: (1) trials exhibiting unusual temporal compression of less than 10 seconds for total duration judgment, (2) trials where any single subevent duration exceeded the total duration, indicating possible data entry errors, and (3) trials with a response time of less than 0.5 seconds. Excluding these trials ensured a more precise and reliable analysis of the remaining data.

The analysis of drawings

To assess participants' recall accuracy and the details recalled, we invited coders to evaluate participants' drawings based on gist and details. We defined gist as any element indispensable to the interpretation of what occurred in a subevent (regardless of style, subjective feelings, etc.), which cannot be altered or excluded without changing the overall meaning of the subevent. After judging the gist accuracy of all participants' drawings for a particular video event, coders checked the drawn items against a provided checklist. This checklist was compiled by having other coders analyze the videos and list all relevant details, including actions, characters, items, and environments. Items not on the checklist (potential false memory items) were noted, and coders reviewed the video to verify their presence. Six coders were involved, divided into three groups, each composed of two coders, to ensure inter-rater reliability. Each group was responsible for coding 10 videos or events. If the two coders disagreed, a third coder was consulted, and the final coded result was determined by majority decision.

Behavioral data analysis

The statistical analyses and graphical representations were performed using R (version 2023.03.3; R Core Team, 2023) and Prism (version 9.4.1 for Windows). We employed linear mixed-effects models (LMMs) to account for various factors and variances. Fixed factors included different levels of memory hierarchy, such as subevent number, gist accuracy, and the number of details recalled for duration judgment. Random effects accounted for participant and event (video) variability at the trial level. All linear mixed effects models (LMMs) were conducted using the lme4 package for linear mixed effect modeling (Bates et al., 2015). A forward selection method was implemented to compare the AICs and BICs of key hypothetical models. The significance levels for fixed and random effects, as well as all Likelihood ratio tests used to compare model fits, were assessed using the anova function in the 'lmerTest' package (Kuznetsova et al., 2017). Post-hoc analyses were conducted using the 'emtrends' function in the 'emmeans' library to compare the effects of different estimation errors.

Results

Despite spending considerable time on watching multiple videos, participants remained sensitive to the difference in subevent duration on both Day 1 and Day 7. Specifically, participants successfully distinguished between 20-second and 40-second subevents (Day 1, $t(47) = -5.253$, $p < .001$, $d = 0.758$; Day 7, $t(47) = -5.31$, $p < .001$, $d = 0.766$) (Fig. S1). In line with prior research, our findings also reveal significant temporal compression compared to the actual duration of the entire event, starting already on Day 1 ($t(47) = -10.956$, $p < .001$, $d = 1.581$) (Tsao et al., 2022)(Fig. 1e).

Single-Day Duration Judgments Vary by Memory Level

We next focused on the judgment of total event duration within a single day. We wanted to see whether the reported total duration can be determined by the recall of content at different levels of the memory hierarchy. To do this, we employed a linear mixed model (LMM), which allow us to identify the main influencing factors from among random effects that originate from the high variability in reported durations among participants and across different types of events (Lositsky et al., 2016; Safi et al., 2024). We ran the LMM with recalled subevent number, overall gist accuracy, and total recalled details number as fixed factors to predict total event duration judgement. Random effects included participant and event. The analysis revealed that the number of recalled subevents significantly influenced reported total duration. Specifically, for each additional number of subevent recalled, participants' reported durations were, on average, 12.23 seconds longer on Day 1 and 11.06 seconds longer on Day 7 (Day 1: $F(118.251, 1) = 243.531$, $p < .001$, $\eta^2 = 0.67$; Day 7: $F(168.751, 1) = 330.939$, $p < .001$, $\eta^2 = 0.66$). On the other hand, the reported total duration is independent of the order of subevents being recalled within the sequence (Day 1, $F(1,659.25) = 0.88$, $p = .348$, $\eta^2 = 0.001$; Day 7, $F(1,615.33) = 0.79$, $p = .375$, $\eta^2 = 0.001$). Additionally, neither gist accuracy (Day 1, $F(213.144, 1) = 0.230$, $p = .632$, $\eta^2 = 0.001$; Day 7, $F(190.831, 1) = 3.155$, $p = .077$, $\eta^2 = 0.02$) nor the number of details recalled (Day 1, $F(59.279, 1) = 2.232$, $p = .141$, $\eta^2 = 0.04$; Day 7, $F(55.903, 1) = 0.013$, $p = .910$, $\eta^2 < 0.001$) contributed significantly to the reported total duration (Fig. S3). Thus, event duration judgment appears to involve recalling directly the number of subevents with no regard to the temporal order of its subevents nor the retrieval of lower-level memories.

Notably, the duration judgment of an entire event is not merely the sum of all reported subevent duration, despite their significant correlation ($r = .795$, $p < .001$) (see Fig. S4). Specifically, the sum of subevent duration judgments was significantly longer (Day 1: $t(47) = -3.134$, $p = .003$, $d = -0.453$; Day 7: $t(47) = -3.01$, $p = .004$, $d = -0.434$), suggesting that duration judgment of event versus subevent may involve different strategies. To investigate the contribution of lower-order memory to subevent duration judgment, we ran a LMM using subevent gist accuracy and the number of recalled details as fixed factors, with participant and event included as random factors. Indeed, unlike the total duration judgments that depended on memories of number of subevents, judgment of subevent duration relies on lower-level memory structures, including both gist accuracy and the number of details (Gist: Day 1, $F(2,1765.06) = 3.762$, $p = .023$, $\eta^2 = 0.004$; Day 7, $F(2,1712.41) = 3.468$, $p = .031$, $\eta^2 = 0.004$; Details: Day 1, $F(1,1571.27) = 22.520$, $p < .001$, $\eta^2 = 0.01$; Day 7, $F(1,1618.20) = 8.313$, $p = .004$, $\eta^2 = 0.005$) (see Fig. S5).

In essence, event and subevent duration judgments differ mainly in that the former does not draw on lower-level memories, whereas the latter does, suggesting that people adopt different recall strategies depending on the requirement of the duration judgment tasks.

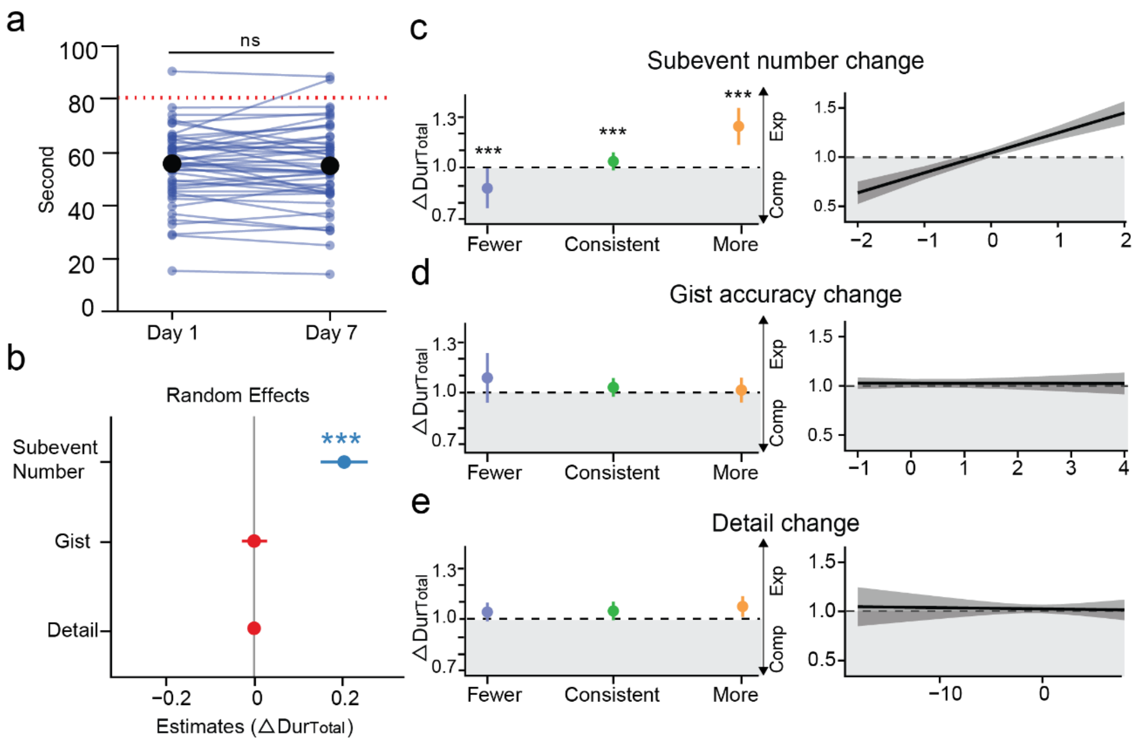

Fig. 2 Event duration changes are only modulated by memory changes of recalled subevent number. (a) Changes in reported event duration among participants across days. Blue dots represent the mean reported duration for each participant. Black dots indicate the mean reported duration for all participants on each day. The red dotted line represents the actual event duration. (b) Main effect of across day change in recalled subevent number ($p < .001$), gist accuracy ($p = .975$) and recalled detail number ($p = .826$) on ΔDur_{Total}. (c-e) Plots of estimated ΔDur_{Total} across days by change of (c) recalled subevent number, (d) gist accuracy and (e) details respectively. Left panel, data were grouped based on the changes of memory across days with Fewer (Blue), Consistent (Green) and More (Orange) from Day 1 to Day 7. Right panel, a continuous estimation of how unit changes on memory affect retrospective duration. The dashed line represents the grey area where $\Delta Dur_{Total} = 1$, indicating no compression. Grey area below 1 represents $\Delta Dur_{Total} < 1$, indicating compression on Day 7. Error bars and shaded areas indicate the 95% confidence intervals (CI).

*** $p < .001$.

Across days: Unique role of event boundary on event duration judgments

After encoding, memory is subjected to transformation over time. By examining how duration judgment may be altered accordingly, these natural memory changes provide a unique opportunity for us to probe how durational information may be related to stored memories. We first confirmed that memory changes occurred over our experimental time window, by comparing Day 1 and Day 7 memory performance using paired sample t-tests. We observed an overall decline in memory performance across all memory levels between Day 1 and Day 7. Specifically, the accuracy of recalled number of subevents ($t(47) = 4.051$, $p < .001$, $d = 0.512$) dropped significantly, with the occurrences of both forgetting (number of occurrences across all participants: 79 out of 593) and false memories (number of occurrences across all participants: 52 out of 593). Moreover, subevent gist accuracy ($t(47) = 4.00$, $p < .001$, $d = 0.411$) and the number of accurately recalled details ($t(47) = 5.842$, $p < .001$, $d = 0.843$) also showed significantly decrease (Fig. 1e).

We next looked at the change in reported total event duration across days, quantified as the ratio of Day 7 to Day 1 (ΔDur_{Total}), with >1 values indicating expanded duration judgments relative to Day 1, a value of 1 indicating consistent duration judgment as Day 1, and <1 indicating more compressed duration judgment relative to Day 1. We employed LMM to test whether this ratio can be explained by changes in recalled subevent number, gist accuracy and number of recalled details across days as fixed factors, and participants and event as random factors. Consistent with the above results for single days, ΔDur_{Total} was significantly affected by changes in recalled subevent number, with temporal compression associated with forgetting whereas temporal expansion linked to falsely inserted subevents ($F(1, 692.32) = 55.34$, $p < .001$, $\eta^2 = 0.07$) (Fig. 2). However, ΔDur_{Total} remained unaffected by changes in overall gist accuracy ($F(1, 491.28) = 0.001$, $p = .975$, $\eta^2 < 0.001$) and the total number of recalled details ($F(1, 669.02) = 0.049$, $p = .826$, $\eta^2 < 0.001$). In fact, the fit of this model does not significantly differ from that of a simpler LMM with only a single fixed factor—change in the number of subevents ($AIC = 353.74$; $BIC = 376.67$; $LR = 0.04$; $p = .98$ over the three-factor model including gist accuracy and recalled detail, see Table S1 for full comparison). The observation that ΔDur_{Total} parallelled the change in the number of subevents recalled provides additional and direct support to our conclusion from single-day analysis that event duration judgment depends on higher-level memory and not lower-level ones.

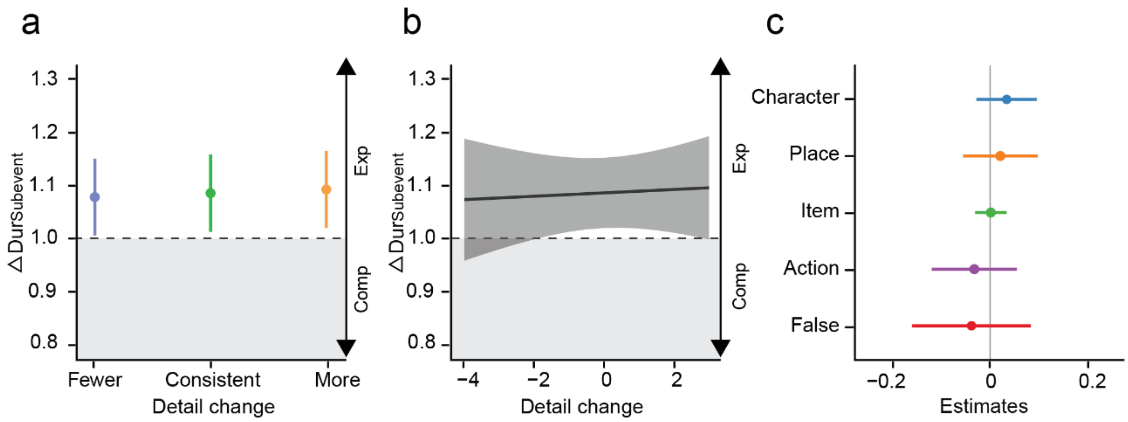

Fig. 3 Subevent duration changes are not modulated by memory changes of recalled detail number. (a) Similar to Fig. 2e but for individual subevent. Fewer (Orange, n = 503 number of subevents), Consistent (Green, n = 497), More (Blue, n = 495). No significant effect is found (*ps* > .05). (b) Continuous estimation. (c) Predicting subevent $\Delta Dur_{Subevent}$ by the change of detail per different subcategories (character, place, item, action, false memory item). across days. No significant main effect was found (*ps* > .05). For the frequency distribution of overall detail changes and by type, see Fig. S7. Error bars and shaded areas indicate the 95% confidence intervals.

Across days: Subevent duration judgment change is dissociated from memory change

Next, we focus on individual subevents and explore similarly whether and how changes in duration judgments are influenced by alterations in recalled subevent-specific memories. As mentioned, participants recalled fewer subevent details over the course of several days ($t(47) = 5.842$, $p < .001$, $d = 0.843$). As participants may have forgotten multiple subevents on Day 7, making it impossible to identify their paired subevent and corresponding duration, we matched subevents across days based on their gist and calculated the $\Delta Dur_{Subevent}$ for each pair (Day 7 / Day 1). Within these matched subevents, we calculated the absolute change in the number of recalled details and conducted a one-sample t-test, revealing a significant change across days, including forgotten details and false memories ($t(47) = 25.35$, $p < .001$, $d = 3.66$). We employed a LMM to test whether the recalled number of details, as a fixed factor, contributed to changes in subevent duration, while accounting for the random effects of participants and events. Despite a general reduction in the number of recalled details, participants' subevent duration judgments did not change significantly ($\beta = -0.01$, $SE = 0.03$, $t(1458.74) = -0.38$, $p = .702$) (Figs. 3a, b). Likewise, duration judgements were not significantly altered when participants remembered a greater number of details on Day 7 relative to Day 1 (β = 0.01, SE = 0.03, t(1475.13) = 0.24, p = .811). Furthermore, when comparing this model to the baseline model, which excluded memory level (i.e., detail recall) as a fixed factor, the results showed no significant improvement. This suggests that memory recall, particularly the recall of details, did not significantly influence judgments of subevent duration (LR = 0.1, p = 0.752).

While our findings suggest that changes in subevent duration judgment are independent of the overall number of details recalled, specific types of details might still have an impact. Action-related details, for instance, may be more significant predictors of subevent duration than static items, as perceived speed influences time perception (Burt, 1999; Burt & Popple, 1996; Mioni et al., 2015; Sasaki et al., 2013). To explore this, we further classified participants' recalled details in our previous LMM into distinct categories, including character, action, item, place, and falsely-recalled details (Fig. S8). However, LMM results shows no significant main effects for any of the categories on subevent $\Delta Dur_{Subevent}$: change in character, $F(1,1474)=1.122$, $p = .290$, $\eta^2 = 0.001$; action, $F(1,1449.8) = 0.274$, $p = .601$, $\eta^2 < 0.001$; item, $F(1,1461.9)= 0.001$, $p = .978$, $\eta^2 < 0.001$; place, $F(1,1462.9)=0.292$, $p = .589$, $\eta^2 < 0.001$, and; falsely recalled item, $F(1,1452.8) = 0.118$, $p = .731$, $\eta^2 < 0.001$ (see Fig. 3c and Table S2).

Averaging tendency of subevent duration judgments occurred across days

If there is an overall decrease in recall detail across days, could this suggest the formation of a memory schema, resulting in a generalization of duration judgements, thus explaining the minimal effect of recalled subevent content on the change of subevent judgements above? To test this, we investigated whether participants tended to perceive each subevent duration more uniformly on Day 7 than on Day 1. We calculated the deviation of each participant's reported subevent duration from their average reported subevent duration for each day, and then analyzed how these deviations changed across all subevents between Day 1 and Day 7. Interestingly, we observed that participants' reported subevent durations tended to converge towards their average over days ($t(47) = 5.32$, $p < .001$, $d = 0.768$) (Fig. S9). Moreover, there is a significant decrease in the difference in duration judgments between 20s- and 40s-subevents across days, as revealed by a paired t-test ($t(47) = -2.02$, $p = .04$, $d = 0.291$). All these suggest that the uniqueness of the subevent duration experience tend to decrease as time passes.

Discussion

Human experience is continuous, yet it can be structurally segmented into meaningful events and further into subevents during encoding and retrieval. We investigated how changes in hierarchical memory components, including event boundaries, gist, and details of subevent, influence retrospective duration judgment. We employed a within-subject test-retest design, enabling the simultaneous measurement of both memory changes and duration judgments. To ensure a well-defined event structure, we used video clips with sharp boundaries, validated by participants' key presses during boundary detection. Our findings indicate that first, people's judgments of event duration rely primarily on higher-level memories, that is their recall of the number of subevents. Second, the judgment of subevent duration is dependent on lower-level memory processes, including the accuracy of gist recall and the number of details remembered, with a tendency towards averaging across days.

Our results suggest that individuals' judgments of total event duration involve recalling higher-level event boundaries of the experience and retrieving associated duration information directly from each segment without recalling lower-level content. This may be achieved through a memory search mechanism where event boundaries act as anchor points that may also induce our direct recall of the associated durational information (Michelmann et al., 2023). Importantly, this duration information may not necessarily be stored as exact numerical values but could be in the form of temporal schemas that capture similarities and differences in event durations and the binding of sequential information, which can later be translated into numerical representations of duration during duration judgments (Baldassano et al., 2018). In this process, finer details, such as the gist accuracy of subevents and the number of details recalled, may be neglected and therefore do not significantly impact our duration judgment.

Regarding duration judgments of subevent, we found that the underlying processes may differ from those used for entire events. Specifically, subevent duration judgments rely on lower-level memories, despite their relatively small effect. This distinction is further highlighted by our across-day analyses, which showed that changes in higher-level memories influenced event duration judgments, whereas alterations in lower-level memories did not significantly impact subevent duration judgments. This finding suggests that the event segmentation model (Block, 1992; Ornstein, 1969; Zakay & Block, 2004), which posits that people perceive duration retrospectively by recalling discrete segments marked by significant changes in context, activity, or goals, may not universally apply to subevents that contains no further segments. Nevertheless, while event duration judgments remained consistent across Day 1 and Day 7, subevent duration judgments tended to average over time. We hypothesized that this averaging tendency may be due to increased similarity among subevents across different events as experiences within similar contexts often result in analogous representations, thus diminishing the distinctiveness of each video's context over time and facilitated generalization of durational schemas (Baldassano et al., 2018; Reagh & Ranganath, 2023). Future studies can explore how memory schematization may affect duration judgement.

These findings prompt a reflection on whether duration judgment is a reconstructive process based on memory or a direct recall of temporal information. Some researchers suggest that duration judgment is influenced by the volume of retrievable memory content, particularly by boundaries marked by contextual changes, as described by the storage size and event segmentation models (Block, 1992; Ornstein, 1969; Zakay & Block, 2004). In contrast, other studies argue that duration is directly encoded and functions independently of memory accuracy. For instance, individuals have been found to accurately judge the duration of past public events even when their memories are imprecise (Burt, 1992; Burt & Kemp, 1991; Friedman, 1993; Yarmey, 2000). Our study suggests that the approach to judging duration is mixed and contingent on the event's structure and complexity. For larger, more complex events, we rely on higher-level memories such as boundaries, whereas for subevents, we rely on lower-level memories. Future research could specifically investigate how modifying event boundaries after initial encoding may change the duration judgement.

In conclusion, people’s judgments of duration vary depending on the complexity and structure of events. Events with subevents rely on higher-order memories, while subevent depend on lower-order memories. Preserving the accuracy of event durations is crucial, as it shapes our perception of the past and supports precise and adaptable time estimations in various contexts. Indeed, individuals with memory impairments, such as those caused by brain injury or neurodegenerative conditions like Alzheimer’s disease, often experience distorted time perception (El Haj et al., 2013; El Haj & Kapogiannis, 2016; Mioni et al., 2014). These temporal inaccuracies not only disrupt recollections of the past but may also affect how we perceive ourselves, i.e., self-identity (Piras et al., 2014). However, our data suggest a novel perspective, indicating that time perception may be more closely linked to the structure of events rather than solely depending on memory accuracy. Therefore, enhancing memory of event structure after encoding could actively preserve duration perception. Future research should explore whether strategies focusing on event structure might help maintain subjective retrospective duration among healthy and among clinical population.

Author Contributions

W.W.Y.Y: Conceptualization, Investigation, Methodology, Resources, Software, Formal Analysis, Visualization, Writing – Original draft, Writing – Review & Editing. J.L.: Methodology, Writing – Review & Editing. Z.Y.: Validation, Writing - Review & Editing. Y.Z.: Writing – Review & Editing. Z.M.: Writing – Review & Editing. X.H.: Conceptualization, Methodology, Funding Acquisition, Supervision, Writing – Original draft, Writing – Review & Editing.

Acknowledgements

We would like to thank Dr. Danni Chen for methodological suggestions, as well as Eunice Choi, Antonius Tam and Sunwoo Joo for assisting with picture coding.

Fundings

The research was supported by the Ministry of Science and Technology of China STI2030-Major Projects (No. 2022ZD0214100), National Natural Science Foundation of China (No. 32171056), General Research Fund (No. 17614922) of Hong Kong Research Grants Council to X. H.

Conflicts of Interest

All authors have no conflicting interests to declare.

Reference

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, *38*(45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

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*(3), 1459–1472. https://doi.org/10.3758/s13414-019-01829-x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Ben-Yakov, A., Eshel, N., & Dudai, Y. (2013). Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *Journal of Experimental Psychology. General*, *142*(4), 1255–1263. https://doi.org/10.1037/a0033558

Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, Action and Cognition: Towards Bridging the Gap* (pp. 141–152). Springer Netherlands. https://doi.org/10.1007/978-94-017-3536-0_16

Burt, C. D. B. (1992). Reconstruction of the duration of autobiographical events. *Memory & Cognition*, *20*(2), 124–132. https://doi.org/10.3758/BF03197160

Burt, C. D. B. (1999). Categorisation of action speed and estimated event duration. *Memory (Hove, England)*, *7*(3), 345–355. https://doi.org/10.1080/096582199387968

Burt, C. D. B., & Kemp, S. (1991). Retrospective duration estimation of public events. *Memory & Cognition*, *19*(3), 252–262. https://doi.org/10.3758/BF03211149

Burt, C. D. B., & Popple, J. S. (1996). Effects of implied action speed on estimation of event duration. *Applied Cognitive Psychology*, *10*(1), 53–63. https://doi.org/10.1002/(SICI)1099-0720(199602)10:1<53::AID-ACP362>3.0.CO;2-S

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*, *29*(3), 162–183. https://doi.org/10.1002/hipo.23074

D’Argembeau, A., Jeunehomme, O., & Stawarczyk, D. (2022). Slices of the past: How events are temporally compressed in episodic memory. *Memory*, *30*(1), 43–48. https://doi.org/10.1080/09658211.2021.1896737

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, *88*(1), 20–32. https://doi.org/10.1016/j.neuron.2015.09.004

El Haj, M., & Kapogiannis, D. (2016). Time distortions in Alzheimer’s disease: A systematic review and theoretical integration. *Npj Aging and Mechanisms of Disease*, *2*(1), 1–5. https://doi.org/10.1038/npjamd.2016.16

El Haj, M., Moroni, C., Samson, S., Fasotti, L., & Allain, P. (2013). Prospective and retrospective time perception are related to mental time travel: Evidence from Alzheimer’s disease. *Brain and Cognition*, *83*(1), 45–51. https://doi.org/10.1016/j.bandc.2013.06.008

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*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, *113*(1), 44–66. https://doi.org/10.1037/0033-2909.113.1.44

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*(6), 457–467. https://doi.org/10.1101/lm.550407

Gransier, R., Peeters, S., & Wouters, J. (2023). The importance of temporal-fine structure to perceive time-compressed speech with and without the restoration of the syllabic rhythm. *Scientific Reports*, *13*(1), 2874. https://doi.org/10.1038/s41598-023-29755-x

Grondin, S., Laflamme, V., & Désautels, F. (2014). The delay before recall changes the remembered duration of 15-minute video sequences. *Applied Cognitive Psychology*, *28*(5), 677–684. https://doi.org/10.1002/acp.3053

Jeunehomme, O., & D'Argembeau, A. (2019). The time to remember: Temporal compression and duration judgements in memory for real-life events. *Quarterly Journal of Experimental Psychology*, *72*(4), 930–942. https://doi.org/10.1177/1747021818773082

Jeunehomme, O., Folville, A., Stawarczyk, D., Van Der Linden, M., & D'Argembeau, A. (2018). Temporal compression in episodic memory for real-life events. *Memory*, *26*(6), 759–770. https://doi.org/10.1080/09658211.2017.1406120

Karmarkar, U. R., & Buonomano, D. V. (2007). Telling time in the absence of clocks. *Neuron*, *53*(3), 427–438. https://doi.org/10.1016/j.neuron.2007.01.006

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Little, D. R., Oehmen, R., Dunn, J., Hird, K., & Kirsner, K. (2013). Fluency profiling system: An automated system for analyzing the temporal properties of speech. *Behavior Research Methods*, *45*(1), 191–202. https://doi.org/10.3758/s13428-012-0222-0

Lositsky, O., Chen, J., Toker, D., Honey, C. J., Shvartsman, M., Poppenk, J. L., Hasson, U., & Norman, K. A. (2016). Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *eLife*, *5*, e16070. https://doi.org/10.7554/eLife.16070

MacDonald, C. J. (2014). Prospective and retrospective duration memory in the hippocampus: Is time in the foreground or background? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1637), 20120463. https://doi.org/10.1098/rstb.2012.0463

Michelmann, S., Hasson, U., & Norman, K. A. (2023). Evidence that event boundaries are access points for memory retrieval. *Psychological Science*, *34*(3), 326–344. https://doi.org/10.1177/09567976221128206

Mioni, G., Grondin, S., & Stablum, F. (2014). Temporal dysfunction in traumatic brain injury patients: Primary or secondary impairment? *Frontiers in Human Neuroscience*, *8*, 269. https://doi.org/10.3389/fnhum.2014.00269

Mioni, G., Zakay, D., & Grondin, S. (2015). Faster is briefer: The symbolic meaning of speed influences time perception. *Psychonomic Bulletin & Review*, *22*(5), 1285–1291. https://doi.org/10.3758/s13423-015-0815-6

Ornstein, R. E. (1969). *On the Experience of Time*. Harmondsworth.

Paton, J. J., & Buonomano, D. V. (2018). The neural basis of timing: Distributed mechanisms for diverse functions. *Neuron*, *98*(4), 687–705. https://doi.org/10.1016/j.neuron.2018.03.045

Piras, F., Piras, F., Ciullo, V., Danese, E., Caltagirone, C., & Spalletta, G. (2014). Time dysperception perspective for acquired brain injury. *Frontiers in Neurology*, *4*. https://www.frontiersin.org/articles/10.3389/fneur.2013.00217

Pu, Y., Kong, X.-Z., Ranganath, C., & Melloni, L. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature Communications*, *13*, 622. https://doi.org/10.1038/s41467-022-28216-9

Radvansky, G. A., Doolen, A. C., Pettijohn, K. A., & Ritchey, M. (2022). A new look at memory retention and forgetting. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *48*(11), 1698–1723. https://doi.org/10.1037/xlm0001110

Reagh, Z. M., & Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nature Communications*, *14*(1), 1279. https://doi.org/10.1038/s41467-023-36805-5

Safi, O. K., Shi, Y., Madan, C. R., Lin, T., & Palombo, D. J. (2024). The effects of emotion on retrospective duration memory using virtual reality. *Psychological Research*, *88*(3), 974–986. https://doi.org/10.1007/s00426-023-01909-6

Santoro, A., Frankland, P. W., & Richards, B. A. (2016). Memory transformation enhances reinforcement learning in dynamic environments. *The Journal of Neuroscience*, *36*(48), 12228–12242. https://doi.org/10.1523/JNEUROSCI.0763-16.2016

Sasaki, K., Yamamoto, K., & Miura, K. (2013). The difference in speed sequence influences perceived duration. *Perception*, *42*(2), 198–207. https://doi.org/10.1068/p7241

Sun, C., Yang, W., Martin, J., & Tonegawa, S. (2020). Hippocampal neurons represent events as transferable units of experience. *Nature Neuroscience*, *23*(5), 651–663. https://doi.org/10.1038/s41593-020-0614-x

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*(2), 236. https://doi.org/10.1037/a0015631

Terada, S., Sakurai, Y., Nakahara, H., & Fujisawa, S. (2017). Temporal and rate coding for discrete event sequences in the hippocampus. *Neuron*, *94*(6), 1248-1262.e4. https://doi.org/10.1016/j.neuron.2017.05.024

Tsao, A., Yousefzadeh, S. A., Meck, W. H., Moser, M.-B., & Moser, E. I. (2022). The neural bases for timing of durations. *Nature Reviews. Neuroscience*, *23*(11), 646–665. https://doi.org/10.1038/s41583-022-00623-3

Wallenstein, G. V., Eichenbaum, H., & Hasselmo, M. E. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences*, *21*(8), 317–323. https://doi.org/10.1016/s0166-2236(97)01220-4

Yarmey, A. D. (2000). Retrospective duration estimations for variant and invariant events in field situations. *Applied Cognitive Psychology*, *14*(1), 45–57. https://doi.org/10.1002/(SICI)1099-0720(200001)14:1<45::AID-ACP623>3.0.CO;2-U

Yarmey, A. D., & Matthys, E. (1990). Retrospective duration estimates of an abductor's speech. *Bulletin of the Psychonomic Society*, *28*(3), 231–234. https://doi.org/10.3758/BF03334012

Zakay, D., & Block, R. A. (2004). Prospective and retrospective duration judgments: An executive-control perspective. *Acta Neurobiologiae Experimentalis*, *64*(3), 319–328. https://doi.org/10.55782/ane-2004-1516

Zakay, D., & Fallach, E. (1984). Immediate and remote time estimation—A comparison. *Acta Psychologica*, *57*(1), 69–81. https://doi.org/10.1016/0001-6918(84)90054-4

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language*, *122*(3), 151–161. https://doi.org/10.1016/j.bandl.2011.12.010

Supplementary Materials

Retrospective duration judgments of naturalistic events depend on memories of event boundaries

Winny W. Y. Yue, Jing Liu, Ziqing Yao,
Yuqi Zhang, Zexuan Mu, and Xiaoqing Hu

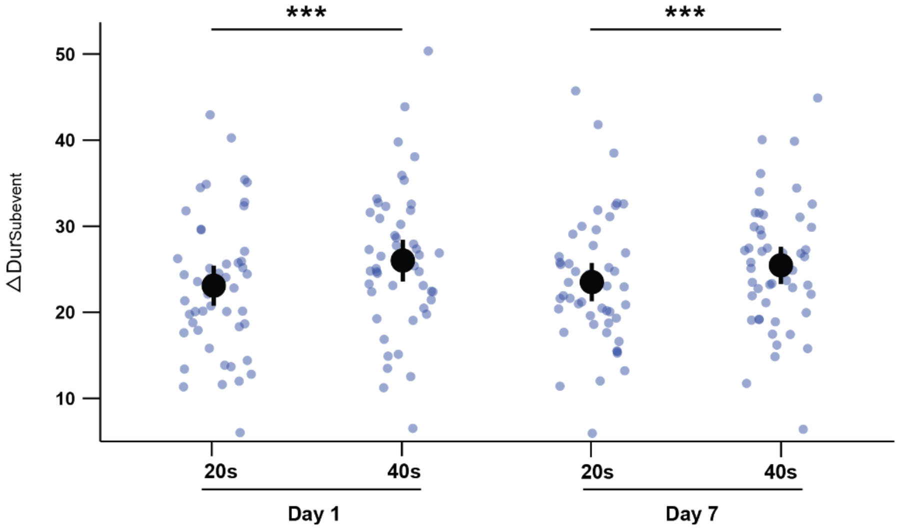

Fig. S1. Reported duration for 20-s and 40-s subevent on both Day 1 and Day 7. Significant differences were found on between 20-s and 40-s subevent on Day 1 ($t(47) = -5.253$, $p < .001$, $d = 0.758$) and on Day 7 ($t(47) = -5.31$, $p < .001$, $d = 0.766$). Blue dots represent the mean reported duration for each participant. Black dots indicate the mean reported duration for all participants on each day. Error bars indicate the 95% confidence intervals. Statistical significance is denoted by stars, with *** indicating $p < 0.001$.

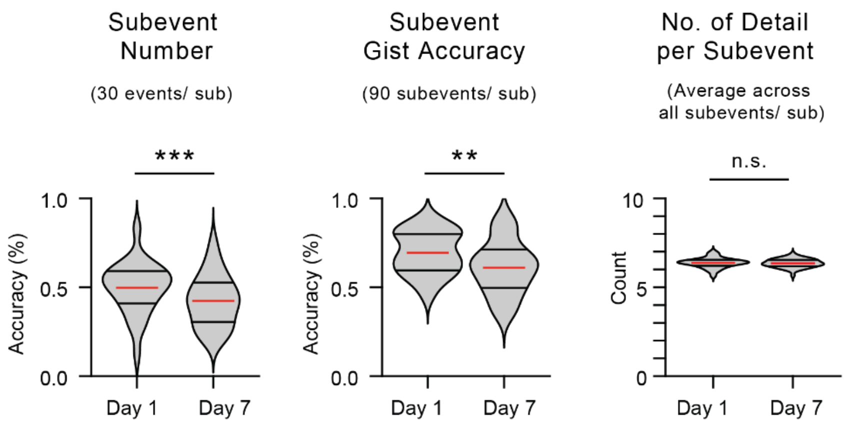

Fig. S2. Memory performance across days. Left, for the reported subevent number. Middle, for the recalled gist accuracy. Right, for the number of details. Significant differences were found for changes in the number of subevents ($t(47) = 5.232$, $p < .001$, $d = 0.755$) and subevent gist accuracy ($t(47) = 6.051$, $p < .001$, $d = 0.601$), indicating forgetting. However, no significant difference was found for the number of details per subevent across days ($t(47) = 0.475$, $p = .637$, $d = 0.073$). Statistical significance is denoted by stars, with ** indicating $p < 0.01$, and *** indicating $p < 0.001$.

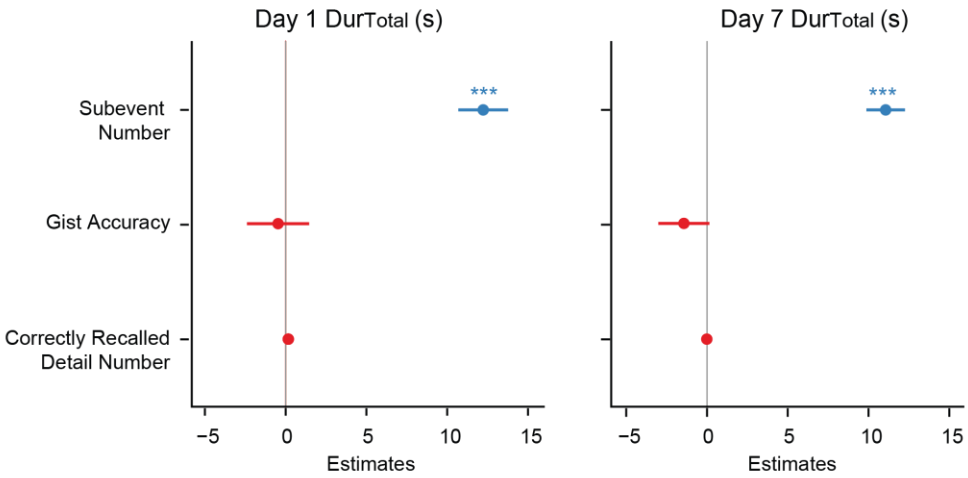

Fig. S3. LMM results on Day 1 and Day 7 reported Dur_{Total} with including subevent number, gist accuracy, and correctly recalled detail number, and random effects for participant and event. Only the number of recalled subevents significantly predicts the Dur_{Total} on both Day 1 ($F(1,118.251) = 234.531$, $p <.001$, $\eta^2 = 0.67$) and Day 7 ($F(1,168.751) = 330.939$, $p <.001$, $\eta^2 = 0.66$). Statistical significance is denoted by stars, with *** indicating $p < 0.001$.

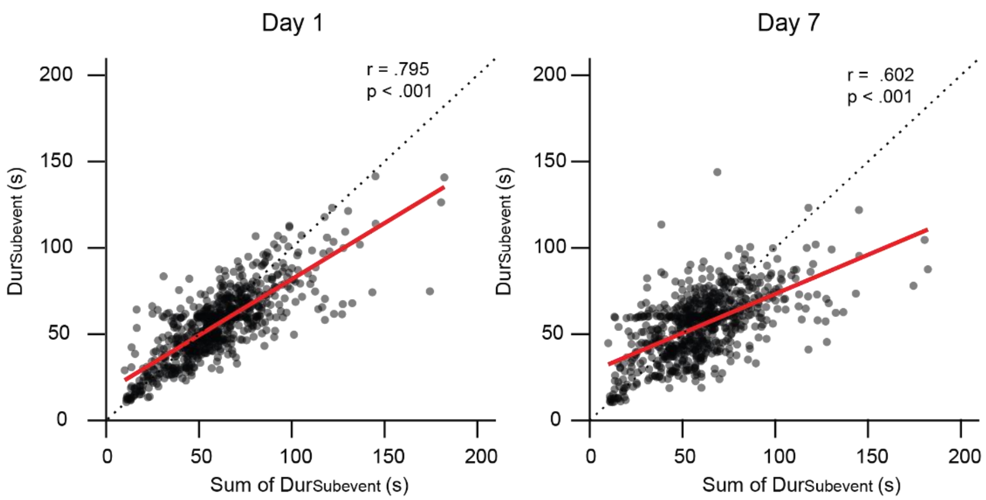

Fig. S4. Significant positive correlations between the sum of all reported $Dur_{Subvent}$ and Dur_{Total} on both Day 1 ($r = .795$, $p <.001$) and Day 7 ($r = .602$, $p <.001$). The red line represents the best-fit regression line, while the black dotted line indicates the diagonal. Each dot represents one video trial.

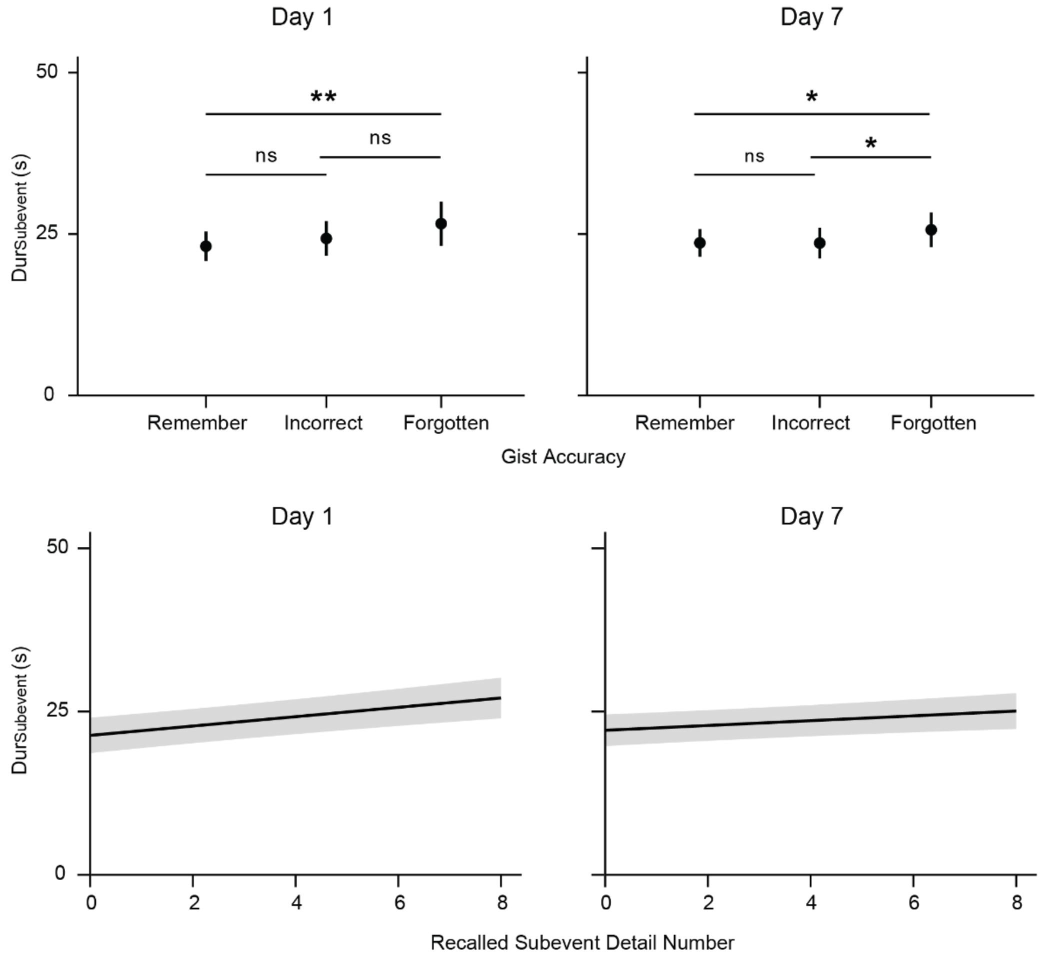

Fig. S5. Predicted $Dur_{Subvent}$ in relation to subevent gist accuracy and subevent detail number on Day 1 and Day 7. Linear mixed-effects model (LMM) results for $Dur_{Subvent}$ include subevent gist accuracy, original duration, correctly recalled detail number, and day as fixed factors, with random effects for participants and events. Both gist accuracy and detail number were significant predictors of $Dur_{Subvent}$ on Day 1 ($F(2,1765.06) = 3.762$, $p = .023$, $\eta^2 = 0.004$ and $F(1,1571.27) = 22.520$, $p < .001$, $\eta^2 = 0.01$ respectively) and on Day 7 ($F(2,1712.41) = 3.468$, $p = .031$, $\eta^2 = 0.004$ and $F(1,1618.20) = 8.313$, $p = .004$, $\eta^2 = 0.005$ respectively). Error bars and shaded areas indicate the 95% confidence intervals. Statistical significance is denoted by stars, with * indicating $p < 0.05$ and ** indicating $p < 0.01$.

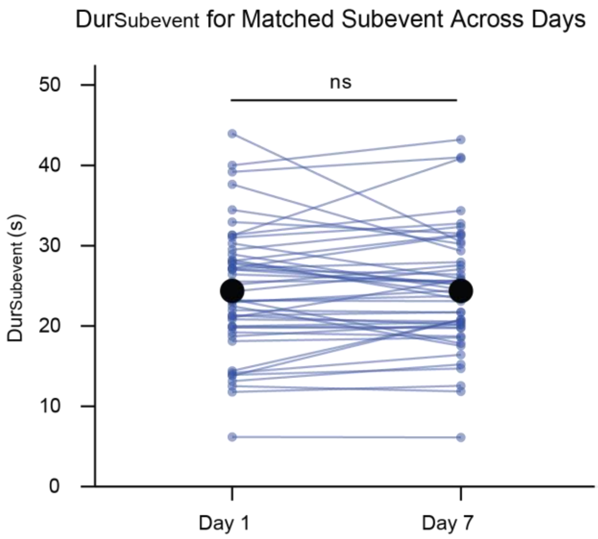

Fig. S6. $Dur_{Subvent}$ on Day 1 and Day 7 by participants for matched subevents in Fig. 3. A paired t-test indicated insignificant differences between the two days ($t(47) = -0.002$, $p = .999$, $d < 0.001$). Blue dots represent the mean $Dur_{Subvent}$ for each participant, while black dots indicate the overall mean $Dur_{Subvent}$ across all participants for each day.

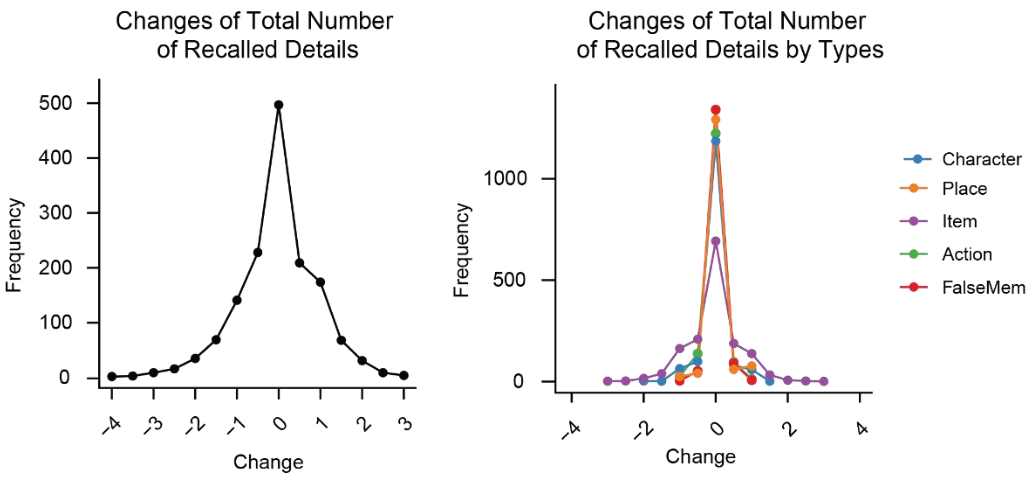

Fig. S7. Left panel, the distribution of detail number change in Fig. 3. Right panel, the distribution of the number of detail changes, color matched with the detail types in Fig. 3c) ($n = 1495$).

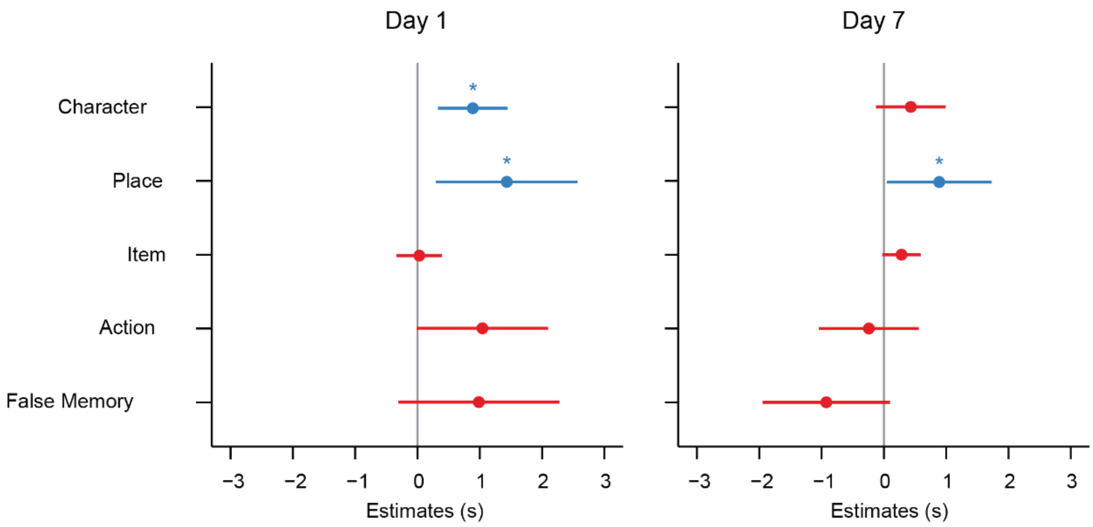

Fig. S8. The main effect of different types of detail on $Dur_{Subvent}$. LMM performed on all $Dur_{Subvent}$ with including subevent gist accuracy, original duration, different types of detail recalled and day as fixed factors, and random effects for participant and event. Subevent durations were significantly predicted by details related to characters ($F(1, 1440.4) = 6.364$, $p = .012$, $\eta^2 = 0.004$) and places ($F(1, 1673.4) = 6.121$, $p = .013$, $\eta^2 = 0.004$) on Day 1, and by places ($F(1, 1667.2) = 4.303$, $p = .038$, $\eta^2 = 0.003$) on Day 7. Statistical significance is denoted by stars, with * indicating $p < 0.05$.

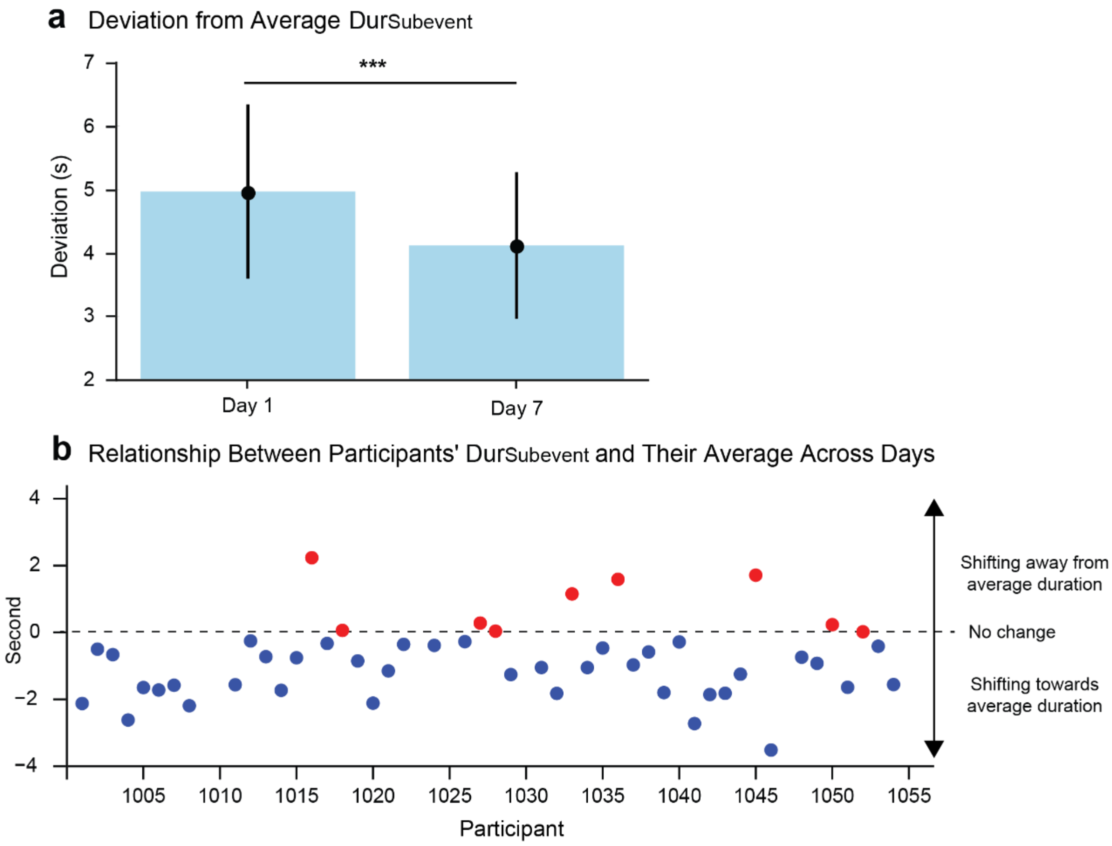

Fig. S9. Plots of participants' reported subevent duration deviated from the average. The average duration was calculated for all subevents across all events for each participant. (a) Participants' reported subevent durations showed a significant decrease in deviation, converging towards their average over the days ($t(47) = 5.32$, $p < .001$, $d = 0.768$). (b) Individual participants' changes in mean deviation. Dots colored in blue are below the dotted line of zero, indicating a tendency shifting toward the mean on Day 7. Dots colored in red represent subjects with a tendency shifting away from the mean. Error bars and shaded areas indicate the 95% confidence intervals. Statistical significance is denoted by stars, with *** indicating $p < 0.001$.

Model	Method	Subevent No. Change	Content Accuracy Change	No. of Detail Change	Random factor: Participant	Random factor: Event	AIC	BIC	BF01
1	LM	V					411.214	424.968	
2	LMM	V			V	V	365.173	388.097	
3	LMM	V	V		V	V	373.864	401.372	0.0005
4	LMM	V		V	V	V	375.710	403.218	0.001
5	LMM	V	V	V	V	V	384.333	416.426	<.001

Table S1. Comparison of results predicting duration of the whole event using different methods. Tick marks indicate the variables included in the model prediction. All linear mixed-effects model (LMM) analyses include participants and events as random factors. Model fits are evaluated using AIC and BIC. The AIC and BIC indicate better model fits for LMM analysis compared to LM analysis. The model comparison results show that the linear model with changes in gist accuracy and detail (Model 5) provided a worse fit relative to the best-fitting model based on BIC (Model 2; *BF01* = 7.052e-07). *BF01* indicates the evidence in favor of the best-fitting model (Model 2) compared to Model 3,4,5.

Model	Method	No. of Detail Change	Character detail change	Action detail change	Action detail change	Place detail change	False detail change	Random factor: Participant	Random factor: Event	AIC	BIC
1	LM							V	V	1717.765	1739.005
2	LMM	V						V	V	1726.7	1753.249
3	LMM		V	V	V	V	V	V	V	1750.33	1798.119

Table S2. Comparison of Results Predicting Duration of the Subevent Using Different Methods. Tick marks indicate the variables included in the model prediction. All linear mixed-effects model (LMM) analyses include participants and events as random factors. Model fits are evaluated using AIC and BIC. The AIC and BIC values indicate model fit, with lower values suggesting better fit. Model 1 demonstrates the best fit based on AIC and BIC values, indicating that the number of detail changes is not a significant predictor for subevent duration changes across days.