

In search of lost time: Reconstructing the unfolding of events from memory



Myrthe Faber, Silvia P. Gennari*

Department of Psychology, University of York, UK

ARTICLE INFO

Article history:

Received 8 March 2015

Revised 24 June 2015

Accepted 25 June 2015

Keywords:

Event memory
Event concepts
Event schemas
Similarity
Event segmentation
Learning
Episodic memory

ABSTRACT

When remembering an event, not only do we recollect what happened, when and where it happened, but also how it unfolded over time. What aspects of events are encoded in memory to support this recollection? This question is central for understanding the nature of event memories and our reconstruction of the time passed. In this article, we investigate how the spontaneous encoding of unfamiliar animations during learning influences the recollection of how these animations unfold. Specifically, we examine two structural properties of dynamic event sequences known to modulate the amount of information encoded in memory: the perceived number of sub-events and their perceived similarity. We found that despite clock duration remaining constant, more sub-events and less similar ones led to longer recognition memory latencies, duration judgments and mental event replaying. In particular, across stimulus animations, both the perceived number of sub-events and their degree of similarity contributed to the prediction of duration judgments and the length of mental event reproductions. Results indicate that the number and nature of sub-events in a sequence modulate how we reconstruct its duration and temporal unfolding, thus suggesting that these event properties, which mediate the amount of information encoded for an event, modulate the subsequent recollection of its temporal unfolding.

© 2015 Elsevier B.V. All rights reserved.

We represent events as unfolding over time. If we have experienced an event a few times—for example a video clip of someone performing dance moves—we can later mentally replay what happened and judge its duration relative to other events in memory. The more we experience different instances of an event, the more we are able to talk and reason about it and to physically or mentally reproduce it. This ability is critical in learning and storing information about the world, and in recounting or evaluating learned information. Thus, how do we reconstruct the unfolding of an event from memory? And specifically, after repeated exposure, what event properties are retained when recalling it?

Previous research in event perception and memory suggests that people's event representations are mediated by the perceptual analysis performed during encoding (Kurby & Zacks, 2008). We naturally and consistently segment our experience into units, and the grain of this segmentation during encoding—whether it is in large or fine-grained units—determines the richness of subsequent recall (Hanson & Hirst, 1989; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Segmentation of familiar experiences (e.g.,

cooking) is guided by previously learned event schemas stored in long-term semantic memory, as we already know what to expect. However, the segmentation of unfamiliar events, like those used in the experiments reported here, relies more heavily on bottom-up (sensory) and subsequent pattern recognition cues. For example, an individual may initially rely on a sensory change (e.g., changes in motion path (Zacks, 2004)) and eventually, over repeated exposures, may recognize segments of experience that tend to co-occur or pattern together in different contexts (Fiser & Aslin, 2002; Gómez & Gerken, 2000; Orbán, Fiser, Aslin, & Lengyel, 2008). Indeed, statistical learning studies using unfamiliar stimuli (syllables, shapes) have shown that babies and adults alike automatically track such regularities in the absence of previous relevant knowledge.

Much of this research however has been focused on *what* we remember of events and *how we update and store* information as we face the flow of experiences (Radvansky & Copeland, 2006; Zacks, Speer, & Reynolds, 2009; Zacks et al., 2007; Zwaan & Radvansky, 1998). Here, we ask whether the organization of unfamiliar events into naturally occurring units also modulates the episodic memory representation of *how they unfold* and *how long* we remember them to be. In particular, we examine how perceived structural characteristics of dynamic events—sub-event

* Corresponding author at: Department of Psychology, University of York, Heslington, York YO10 5DD, UK.

E-mail address: Silvia.gennari@york.ac.uk (S.P. Gennari).

similarity and number of sub-events—modulate memory representations. These structural characteristics are known to operate in memory encoding, learning and retrieval (Miller, 1956; Sloutsky, 2003). Indeed, the more segments a person can find in a sequence of events—e.g., in a film—the more information that person can store and subsequently retrieve about that sequence (Boltz, 1992; Hanson & Hirst, 1989; Zacks et al., 2007). Likewise, similarity between events, and particularly, the repetition of similar events in variable contexts, leads to more efficient encoding, as shown by statistical learning and memory research (Avrahami & Kareev, 1994; Bellezza & Young, 1989; Brady, Konkle, & Alvarez, 2009; Melton, 1967). Over several exposures, similar or repeating events tend to be chunked into one schema in memory, whereas dissimilar events are likely to be stored in distinct units, resulting in more stored information. Indeed, in a previous study, we found that after learning, events composed of repeating sub-events are recognized faster and judged shorter than events composed of dissimilar sub-events, suggesting that repeating events are encoded into simpler structures (Faber & Gennari, 2015). Taken together, these findings suggest that segmental and similarity structure appears to lead to more stored and recollected information.

The present studies therefore investigate the role of these structural characteristics in retrospective duration judgments and event reconstructions as supported by previously encoded event representations (episodic memory). Since our previous study focused primarily on similarity and examined a small unsystematic item set, we aim to generalize and extend these findings, and more importantly, to assess on an item-by-item bases the relative contribution of the number of perceived segments and their similarity. We hypothesize that both these event properties should independently contribute to the items' encoding and thus, explain orthogonal portions of variance in individual items' responses: The more units or the more dissimilar units we recollect about an event, the longer it is remembered to be and the longer it takes to mentally replay. This is because the more information we store as a function of each segmental and similarity structure, the more we recollect about an event and the longer that event takes to mentally reconstruct from memory. Specifically, we predict that as the perceived number of sub-events increases and the similarity between them decreases, longer event representations should be reconstructed, even when clock duration remains constant.

To test this hypothesis, we asked participants to study the content of novel stimulus animations (i.e., what happens in the

animations) over several exposures for a subsequent memory test. A paired-associate learning paradigm was used such that each animation was paired with a still-frame (the cue frame of Fig. 1) to be used as the cue to retrieve the animation content in subsequent tasks. The stimulus animations depicted unfamiliar events consisting of various geometric shapes moving, changing or causing changes. These animations were grouped into triads of three conditions of identical clock duration (see Fig. 1): the *basic condition* contained a basic sequence; the *numerous condition* comprised of the same basic sequence but with an added consistent change; the *dissimilar condition* was like the numerous condition, but this time the added changes were different in nature. After learning, participants performed a memory task in which they indicated whether a series of still frames belonged to the studied animations. The main purpose of this task was to establish whether participants learned the animations. Finally, participants were given additional surprise tasks designed to probe their mental representation of the events' unfolding. Experiment 1 used a duration judgment task and Experiment 2 a mental replay task. Thus, participants did not know in advance the nature of these tasks and had to base their judgment on whatever information they had encoded during learning. See Fig. 2 for a schematic representation of the tasks and trial structure.

This experimental setup encouraged participants to learn and deeply encode the animations according to perceived properties, rather than specific event schemas already stored in semantic memory. Nevertheless, the event structures investigated here resemble structures or event schemas attested in the real world. Indeed, many ordinary events are made up of similar repetitive sub-events, e.g., walking and hammering, whereas others are made up of different or variable sub-events, e.g., building a house and cooking, and are thus comparable to our basic and numerous conditions on the one hand, and the dissimilar condition on the other hand. This resemblance provides an opportunity to infer how we spontaneously structure unfamiliar dynamic events like those found in the real world in the course of learning. Our approach is thus concerned with the spontaneous event memory representations emerging from learning that are then retrospectively relied upon to reconstruct an event and its duration.

In this respect, the approach contrasts with time perception studies and previous retrospective time studies. In time perception studies, participants are instructed to attend to or monitor stimulus duration, and thus they may attempt to time the stimuli or engage

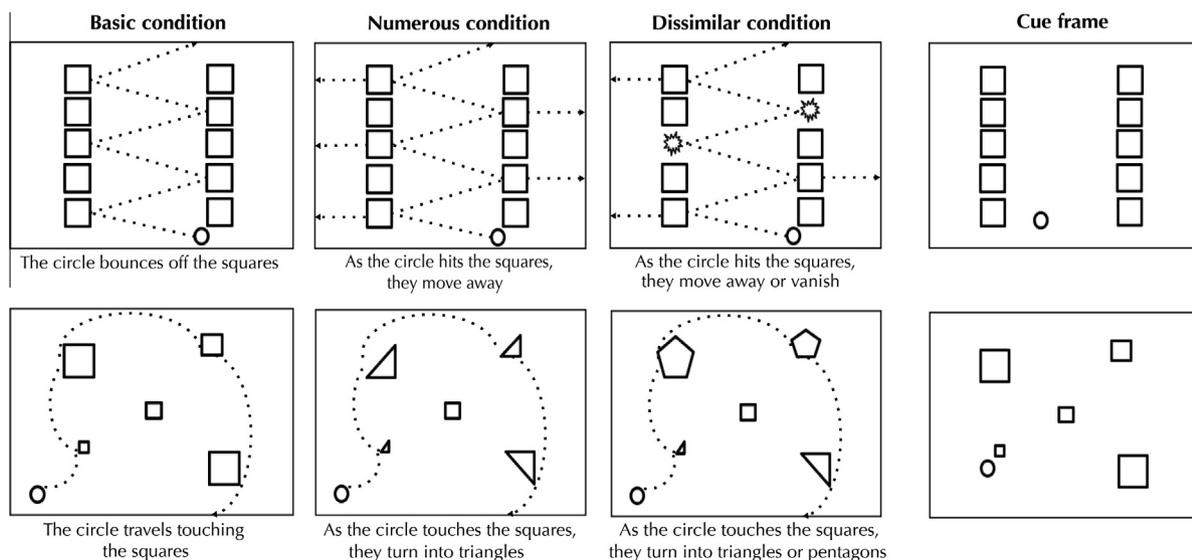


Fig. 1. Example of two stimulus triads (top and bottom rows). Dotted arrows indicate the path of motion. Cue frames were learned together with the animations and were used as a reminder of the animation content in subsequent tests after learning.

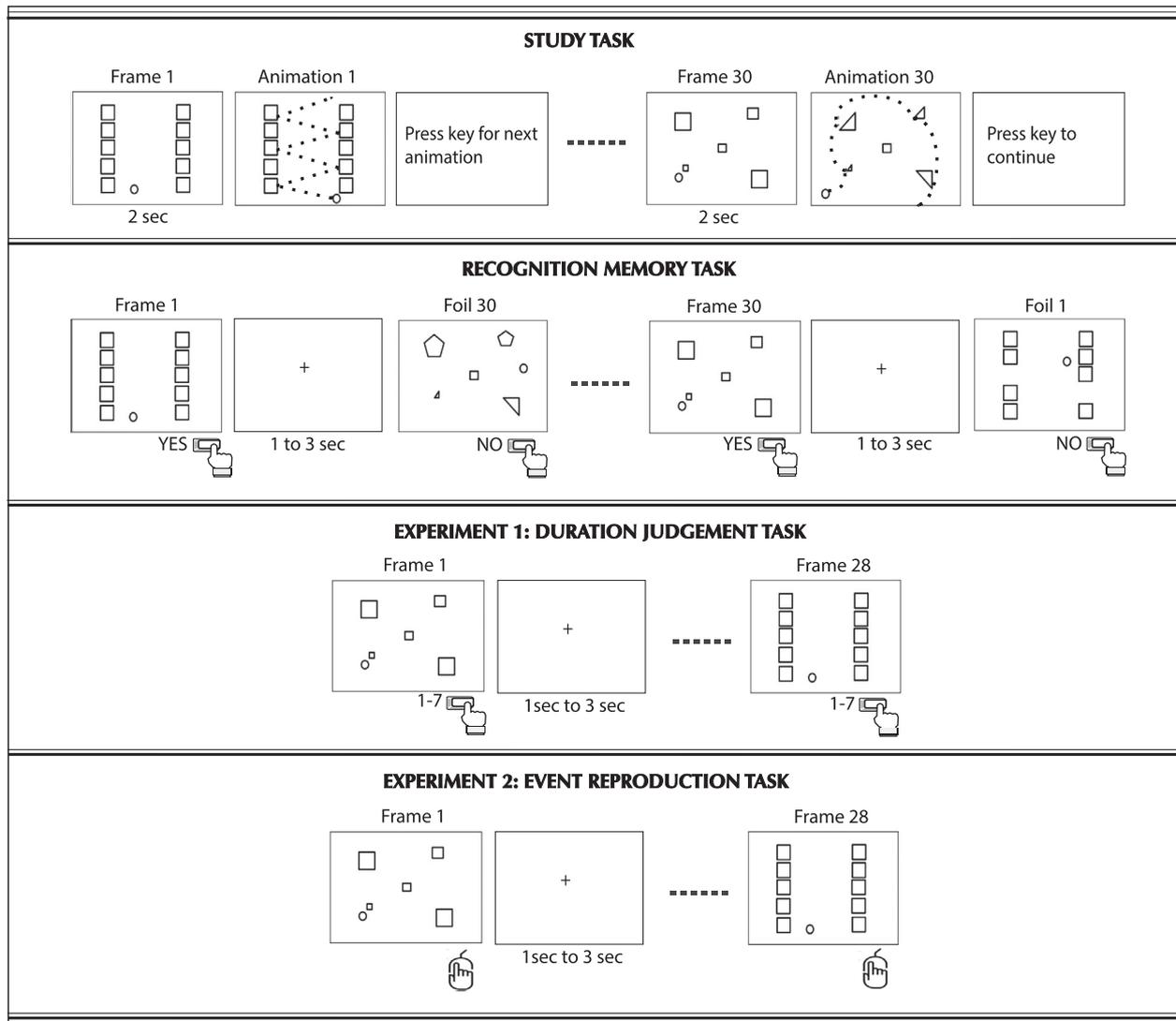


Fig. 2. Schematic representation of the tasks and trial structure in Experiments 1 and 2.

in time-keeping strategies to immediately reproduce duration, judge it numerically or compare it to some other interval (Brown, 1995; Brown & Boltz, 2002; Grondin, 2010; Liverence & Scholl, 2012; Waldum & Sahakyan, 2012). Because of the nature of the task in timing paradigms, they are generally considered to target different cognitive processes from retrospective paradigms, in which participants do not attend to duration but rather reconstruct or infer it from whatever they have attended to during stimulus processing (Block & Zakay, 1997; Grondin, 2001, 2010; Zakay, 1993). All of these time studies attempt to understand how duration is judged or timed from a single stimulus experience, often composed of words or arbitrary stimuli. In contrast, here we examine how equally long novel dynamic events, which vary in segmental and similarity structure, are learned and later reconstructed from episodic memory. We thus aim to contribute to the understanding of episodic event representations, and event cognition more generally.

1. Experiment 1: Recognition memory and retrospective scalar duration judgment

In this experiment, we investigate whether recognition memory and scalar duration judgments are modulated by the perceived

segmental and similarity structure of the remembered stimuli. To this end, we created a set of animations that varied in event structure properties (Fig. 1) and asked a set of independent participants to judge both the number of segments and their similarity. These judgments confirmed our intuitions that the basic and the dissimilar conditions differed in both number of segments and similarity, whereas the numerous condition only differed from the basic one in number of segments, although this difference was relatively small. On the assumption that changes in these event properties lead to more stored information during learning and encoding, and thus lead to more recollected information during duration judgments, we predicted a positive trend across the conditions' means such that the basic and the dissimilar conditions would show the largest difference, with the numerous condition located somewhere in this continuum. To statistically examine the relative role of each event structure property, and in particular, whether they make an independent contribution to performance, we conducted step-wise regression analyses on an item-by-item basis, which captures stimulus individual properties more precisely than our grouping conditions. These analyses allow examining the proportion of variance accounted for at each step over and above other predictors by computing the R^2 change statistics. Thus, we would expect that if say, similarity has a contribution separately

from segmentation, it should significantly increase the variance accounted for by previously entered predictors, such as segmentation.

We conducted two tasks in this experiment after the learning phase. The recognition memory task was primarily conducted to obtain accuracy measures that could be used to exclude poor learners and match conditions, thus reducing the likelihood that the subsequent duration judgments would be contaminated with guesses due to poor learning of one condition over another. However, our previous study has shown that recognition latencies appear to increase for events with dissimilar sub-events (Faber & Gennari, 2015), suggesting that some aspects of the animation content were retrieved during the task. Indeed, this is generally expected in recognition tasks with paired-associate learning and highly similar seen and unseen items, as is the case in our task (Malmberg, 2008; Yonelinas, 2002). Thus, if participants recollect some event information to inform their recognition decisions, a main effect of condition can be expected in the recognition task, with a specific positive increase between the basic and the dissimilar conditions.

The main test of our hypotheses was nevertheless the duration judgment task. In this task, participants were presented with a scale (1 indicating very short durations and 7 representing very long durations in comparison with the other studied animations) and were asked to locate each studied animation on this scale. The shortest and the longest animations in the study set were indicated to participants during instructions so that the task required a relative judgment, i.e., judging the duration of each animation relative to the studied set anchored in the shortest and the longest studied animations. We expected this task to provide information about the nature of the memory representations underlying the duration judgment, because providing this judgment involves retrieving qualitative aspects of the animation content concerning its temporal development.

1.1. Methods

1.1.1. Participants

A total of 83 native English speakers who were students at the University of York were tested (35% males). Of these, eight participants were excluded for poor memory performance. The exclusion criteria aimed to match correct recognition across conditions to minimize the influence of guessing or weak memory content in the duration judgment tasks and to guarantee equal number of participants contributing to each list. Thus, seven participants who had recognition accuracy below 50% in one of the condition or an overall false alarm rate above 50% were excluded (to maintain a counterbalanced number of participants per list, one additional participant with the next worst memory performance for that list was excluded). This resulted in correct recognition performance, the latencies of which were used as dependent measure, being matched across conditions (percentage correct for basic condition (i.e., hits): $M = 90%$, numerous condition: $M = 91%$, dissimilar condition: $M = 87%$; Friedman's test = n.s.).

1.1.2. Stimuli

We created 28 animation triads using Adobe Flash CS5.5, each lasting an integer number of seconds (varying between 3 and 9 s, four animation triads for each of the seven time bins). Triad members had the same clock duration and were arranged into three conditions (Fig. 1): a basic event sequence containing a repeating or stable motion of a shape (*basic condition*), a sequence with a repeating change (e.g., displacement) added onto the basic motion (*numerous condition*), and a sequence like the numerous one but with dissimilar changes (e.g., displacement and disappearance)

(*dissimilar condition*). The basic version in a triad was systematically modified into the numerous version, which in turn was modified into the dissimilar one, keeping speed of motion constant. Across triads, shapes, motion and changes were visually different to prevent memory interference. The stimuli also included, for each triad, a single still frame extracted from near the beginning of the triad animations (one still frame per triad common to all the triad members). These cue frames were used as a retrieval cue in the memory and duration judgment tasks (see Fig. 1). Two additional *anchor animations* (lasting 2 and 10 s respectively) and corresponding cue-frames were also created for use in the duration judgment task.

1.1.2.1. Pre-test studies

In order to examine the distribution of stimulus properties across our grouping categories, separate sets of participants provided similarity and number of sub-event judgments. Two web-based questionnaires were conducted with independent observers. Stimuli were organized in three lists as in the main study (see below). A total of 121 English speakers recruited through the Mechanical Turk completed the questionnaires but four participants were excluded from the segmentation data due to their idiosyncratic scores (they occurred only once in the responses to each animation). A total of 87 participants were used in the segmentation task (29 per list, mean age = 34.6, 52% males) and a total of 30 participants were used in the similarity task (10 per list, mean age = 38.7, 46% males). The web-link provided to participants directed them to a custom-built web page containing a list of our stimulus animations. The *similarity* questionnaire asked participants to judge how similar the events within each animation were to one another in a scale of 1–7 (1 = not similar at all, 7 = very similar). Examples were provided indicating the extreme points of the scale. The animation could be watched as many times as desired by clicking a play button. The order of the animations in the webpage was random. Table 1 shows the mean similarity ratings and sub-event scores for each condition. Animations in the numerous condition were judged to contain similar sub-events to a comparable extent as the sub-events of the basic condition. In contrast, animations in the dissimilar condition were judged to contain less similar sub-events. Repeated measures ANOVA with items as a random factor and similarity rating as dependent variable indicated that there was a main effect of condition ($F(2,54) = 27.05, p < .001; \eta^2 = .50$), and all pair-wise comparisons were highly significant (all p 's < .001) except for that between the basic and the numerous conditions.

The *sub-event* questionnaire used instructions similar to those in segmentation studies (Zacks, Tversky, & Iyer, 2001). Participants indicated the number of instances in which a smallest natural and meaningful unit within the animation finishes and another starts. They were told to watch the animation several times and count these instances. As shown in Table 1, larger number of sub-events were perceived in the numerous and dissimilar conditions compared to the basic condition. Repeated measures ANOVA with items as a random factor and mean number of sub-events as dependent variable indicated a main effect of

Table 1
Mean responses per condition in pre-test studies.

Event properties	Condition		
	Basic	Numerous	Dissimilar
Number of sub-events	4.50 (1.79)	4.73 (1.87)	4.85 (1.85)
Sub-event similarity rating (1–7 scale)	5.06 (1.26)	5.07 (.94)	3.89 (1.14)

Note: Standard deviations are given in parentheses.

condition ($F(2, 54) = 5.53, p = .007; \eta^2 = .17$) with all pairwise comparisons being significant (all p 's $< .05$) except for that between the numerous and the dissimilar conditions. Note that the mean sub-event score differences between the basic and the numerous or dissimilar conditions were relatively small, compared to other segmentation studies. This is because numbers of segments across items varied greatly (e.g., a short three-second animation, however complex, is bound to have fewer sub-events than a nine-second animation). Because of this variability, it is important to examine the relationship between stimulus properties and our dependent variables on an item-by-item basis, as reported later.

1.1.3. Design and procedure

The experimental items were arranged in three lists of 28 animations. Each list contained only one member of each triad (either basic, numerous or dissimilar) but all conditions across triads. Participants were then randomly assigned to one of these lists. In the *study task*, a list of 28 animations plus the two anchor animations (30 animations in total), paired with the corresponding cue-frames, was presented. Each participant was shown his/her assigned list three times in different random orders. Each trial in this task showed the cue-frame for 2 s followed by its associated animation. When the animation finished, a new display prompted for a key press to see the next animation. Participants were instructed to study the content of the animations (i.e., what happens in the animation) with their paired frames, which were to be used in a subsequent memory test as a reminder of the animation content.

In the *recognition memory task*, the 30 cue-frames that a participant had studied plus 30 foil-frames (similar frames to those studied but previously unseen) were presented in random order. Participants indicated whether each frame belonged to one of the studied animations by pressing a YES or NO key. The number of YES and NO responses was equal. A fixation cross was presented on the screen immediately after a participant response was recorded and its duration varied randomly between 1 and 3 s. The foil-frames were extracted from a triad member not studied in the list (e.g., if the basic condition was studied for triad 1, the foil-frame was a middle or late frame from the numerous or dissimilar conditions), thus keeping foils and cue-frames fairly similar and requiring detailed knowledge of the animations.

As indicated above, the proportion of correct recognition was similar across conditions. Overall correct performance was high, with correct recognition and correct rejections averaging 87% across participants. Correct performance for the anchor animations and their foils was also high (mean correct responses across participants = 89%). False alarms averaged 15% across participants. However, compared across conditions, there were more false alarms in the more complex conditions (basic = 2%, numerous = 17%, dissimilar = 27%) suggesting that discrimination was more difficult for these conditions, as there were more opportunities to make mistakes when more features or changes were involved. From this task, we only analyzed response latencies to YES-correct responses. Outlier recognition latencies longer than 3 standard deviations from each participant's mean were excluded from these data (less than 4% of the data) (Ratcliff, 1993). Note that analyses conducted with response times or their log-transforms as dependent variables had an identical pattern of results. Statistics for this and subsequent measures were calculated using standard analysis of variance (repeated measures ANOVA with subjects or items as random factors) because mixed effect models did not provide a convergent solution.

Finally, in the surprise *duration judgment task*, participants were shown the studied cue-frames in random order and were asked to indicate the duration of its associated animation on a 1–7 scale

(1 = *very short*, 7 = *very long*). Instructions indicated (by displaying the anchor animations' cue-frames) that the anchor animations were the shortest and longest in the studied set and where outside of the scale (i.e., equivalent to the 0 and 8th point in the scale). The objective was to indicate that the whole 1–7 scale should be used in providing the judgments. Randomly varied inter-trial times were used as in the recognition task. Note that alternative tasks used in time perception studies were deemed less appropriate because they provide more indirect links to the underlying event memory, e.g., by requiring the translation of the studied event content into clock time (e.g., seconds), or by introducing a "standard" duration for comparison during testing, which would involve comparing different kinds of memory representations both in degree of encoding and their nature. In this respect, our anchor animations, which are learned along the stimulus set, provide better standards for comparison. Given the high memory accuracy across participants, we report results including all ratings except for 38 trials out of a total of 2100 trials in which a participant took longer than 10 s to respond, which were deemed unlikely to represent a confident judgment. Nevertheless, analyses excluding items on the bases of memory performance revealed the same pattern of results (e.g., excluding items whose cue-frames and cue-foils were incorrectly rejected or accepted).

1.2. Results

1.2.1. Memory judgment latencies

We expected that if sub-event and similarity structures are re-activated in this task, a main effect of condition should be observed in response latencies, with positive trends reflecting the re-activation of more event properties across conditions. Mean response latencies for correctly recognized items are shown in Table 2. On average, the items in the numerous condition took 113 ms longer to be correctly recognized than the basic condition, whereas the items in the dissimilar condition took 150 ms longer than the basic condition. Statistics conducted on log-transformed mean response latencies indicated a main effect of condition ($F_1(2, 148) = 6.25, p = .002, \eta^2 = .08; F_2(2, 54) = 4.97, p = .01, \eta^2 = .16$). To test that positive trends across conditions indeed explained most of the variance in the data, as predicted, we conducted planned contrasts for positive trends, which assign the weights $-1, 0$ and $+1$ to the basic, numerous and dissimilar conditions respectively, and then checked that the residual variance of this trend was not significant (Hays, 1981). Positive trends were evident in the data ($F_1(1, 74) = 11.06, p = .001, \eta^2 = .13, F_2(1, 27) = 8.27, p = .008, \eta^2 = .23$) and the residual variance was not significant ($F_1(1, 74) = 1.6, p = \text{n.s.}, F_2(1, 27) = 2, p = \text{n.s.}$). This suggests, as the means indicate, that the numerous condition was placed along the basic-dissimilar continuum, as there was no additional variance to be explained by some other pattern. These results therefore indicate that structural properties attended to during encoding were recollected during recognition memory, with more encoded information leading to increased latencies, consistent with many paired-associate memory studies (Yonelinas, 1999, 2001).

Table 2
Response latencies and duration judgments in Experiment 1.

Task and measure	Condition		
	Basic	Numerous	Dissimilar
Recognition – correct trials (ms)	1466 (31.47)	1579 (33.28)	1616 (28.39)
Duration rating (1–7 score)	3.86 (0.04)	4.03 (0.05)	4.13 (0.05)

Note: Standard errors are provided in parentheses and were computed following Loftus and Mason (1994).

1.2.2. Duration judgment task

We expected that if the number of sub-events and their dissimilarity modulate the memory representation of the events' unfolding—and from this, the duration judgment—the attributed duration should vary across conditions following a positive trend, despite clock duration remaining constant. Table 2 displays the mean ratings for the conditions, which differed significantly ($F_1(2, 148) = 5.54, p = .005, \eta^2 = .07$; $F_2(2, 54) = 3.23, p = .05, \eta^2 = .11$). The dissimilar condition was judged longer than any other condition, and the numerous condition was also judged longer than the basic condition. As in the recognition latencies, the ratings showed positive trends ($F_1(1, 74) = 11.95, p = .001, \eta^2 = .14$; $F_2(1, 27) = 5.99, p = .02, \eta^2 = .18$), with no remaining significant variance in the data ($F_1(1, 74) = 0.23, p = n.s., F_2(1, 27) = .04, p = n.s.$). These findings suggest that event properties encoded in memory are retrieved during duration judgments, with increasing duration attributed to events of the same clock duration as a function of content condition.

1.2.3. Regression analyses

To evaluate the extent of the contribution of sub-event and similarity aspects of the animations on duration judgments, which unlike speeded recognition, involve deeper recollection, we conducted by-item hierarchical multiple regressions and examined the proportion of variance accounted for by sub-event and similarity scores (obtained from the independent pre-test studies). Note that there was no significant correlation between the scores for the number of segments and the similarity scores ($r = -.09, p = .40$). In the first step of the regression model, we included clock duration as a control predictor to account for the systematic variation built across triads (a factor controlled for in the analyses above). This regression model thus contained mean ratings per item as dependent variable and clock duration as predictor. In the second step, we added sub-event scores to the first regression model. This revealed a significant increase in the proportion of variance accounted for: R^2 increased from .47 to .53 ($F_{\text{change}}(1, 81) = 9.35, p = .003$). In the third step, we added similarity ratings to this regression model, which also revealed a significant increase in the proportion of variance accounted for: here R^2 increased from .53 to .57 ($F_{\text{change}}(1, 80) = 7.66, p = .007$). This pattern of significance remained the same regardless of the order in which the predictors were added into the model and regardless of the pre-test measures used—means or medians. Table 3 reports the statistics of the full regression model. Thus, although clock duration is able to explain a significant proportion of duration attributions, both the number of perceived sub-events and sub-event similarity significantly improve the fit, suggesting that participants evaluate duration from the event structure encoded in memory. Specifically, increasing the number of sub-events in a sequence leads to increased duration attributions, whereas increasing the similarity between sub-events leads to decreased

duration judgments (hence, the negative relationship in Table 3), suggesting as hypothesized, that increased dissimilarity relates to increased duration judgments.

1.3. Discussion

Both the recognition latencies for the recognition memory judgments and the duration ratings displayed a significant effect of condition, with a positive trend over conditions. The results of the memory task suggest that correct recognition latencies were modulated by qualitative aspects of the animations. This was likely due to both the nature of the task and the stimuli. Indeed, discrimination between highly similar seen and unseen items makes the task more difficult overall and promotes recollection of episodic details, albeit within the task constraints of rapid decisions (Malmberg, 2008). Moreover, our stimulus conditions were such that the degree of match between a given probe frame and the underlying memory representation of the animations would vary across conditions. The basic condition for example can be represented with a single schema that repeats over time (e.g., same actions, objects and colors occur over time), and therefore, any consistent or inconsistent frame with that basic pattern would be easily detected. In contrast, the dissimilar condition contained changes in colors, paths or shapes (different sub-events). Any potential frame belonging to such an animation would need to be checked against these changing features (the frame could have occurred at earlier or later stages of the animation). This may lead to relatively more features being checked when deciding on complex animations, thus resulting in longer response latencies. Whichever the case, it is clear that qualitative aspects of the animations, which varied as a function of condition, played a role in correct recognition.

Importantly, the duration ratings display strong correlations with an animation's individual properties. In particular, the results showed independent contributions of the number of segments and similarity between them to the duration ratings, over and above clock duration. These findings suggest that during the duration judgment process, animations were recollected and individual properties retrieved, leading to longer duration attributions as the segmental and similarity structures change. This confirms our expectation that qualitative item-based recollection would be involved in the duration judgment task, and furthermore, supports the hypothesis that participants' segmentation and structuring of an animation's content during learning, as suggested by event segmentation theory and the role of similarity in memory (Faber & Gennari, 2015; Kurby & Zacks, 2008), is reflected in scalar duration judgments.

Nevertheless, certain aspects of the judgment task (both here and in our previous study) may not have been ideal. Participants are asked to evaluate the duration of an animation relative to others and place it in an ordinal scale. Because during learning and encoding, participants did not attend to duration but to what happened in the animation, they do not have access to duration information as such to make duration comparisons. The decision processes and perhaps inferences that led to a response may have unduly influenced the results. Indeed, it is possible that decisions result in longer judgments when animations contain more distinctive details relative to others, and thus, estimations might not straightforwardly reflect memory of the unfolding events, but decision processes. Because our ultimate aim is to understand how the temporal development of events is represented, rather than how duration judgments are provided, it is essential to investigate these representations with more straightforward measures. We therefore conducted a similar learning experiment but this time, we probed the content of the memory representations by asking participants to mentally reproduce or replay the events of the

Table 3
Multiple regression model for the rating results of Experiment 1.

	Model	B	SE B	β
1.	Constant	2.14	.23	
	Clock duration	.31	.04	.69*
2.	Constant	1.98	.22	
	Clock duration	.21	.05	.48*
	Number of sub-events	.16	.05	.31*
3.	Constant	2.75	.35	
	Clock duration	.20	.04	.45*
	Number of sub-events	.16	.05	.31*
	Similarity	-.15	.05	-.21*

* Indicates $p \leq .007$.

animation in their original time course. This task is appropriate because it provides a measure of the duration of the mental replaying of the events, without explicitly asking for duration reproduction, as in timing paradigms, and thus directly relies on the encoded content to reconstruct what happened in the events.

2. Experiment 2: Recognition memory and mental reproduction of events

In this experiment, we used the same stimuli and learning paradigm as in Experiment 1. We also used the same recognition memory task to attempt to replicate our previous results. Importantly, to probe the memory representations extracted from the animations, and in particular, the representation of their temporal unfolding, we asked participants to mentally replay the animations exactly as they have experienced them in their original time course. This task was triggered by the presentation of the cue-frame, and participants clicked the mouse when they were done mentally replaying each animation. We measured how long the animations were mentally replayed for. As before, we sought some assurance that participants had indeed learned the animation so as to not contaminate these measures with guesses, using memory scores to exclude participants with poor memory.

2.1. Methods

2.1.1. Participants

A total of 88 native English speakers, students at the University of York, were tested (26% males). Three participants were unable to perform the tasks (noncompliance) and four participants who showed poor memory performance were excluded in order to match recognition memory performance across conditions. The exclusion cutoff was generally set at 50% correct responses in any one of the conditions and/or a false alarm rate of 50% or higher as in Experiment 1, but in order to counterbalance the number of participants per list, one participant with lower accuracy (44% percent correct in one condition) was kept in the dataset. This resulted in non-significant differences across conditions in correct recognition memory (basic: $M = 91\%$, numerous: $M = 90\%$, dissimilar: $M = 90\%$; Friedman's test = n.s.).

2.1.2. Stimuli

The 28 animation triads described in Experiment 1 were also used in this experiment.

2.1.3. Design and procedure

The design of this study was similar to that of Experiment 1. The 28 stimulus triads were arranged into three different lists as was done for Experiment 1 (one condition per triad but all conditions across items). These lists were randomly assigned to participants. The study task was identical to that used in Experiment 1, except that participants were given the option to watch the animations again after the third cycle during learning. This was done to guarantee good memory encoding. Thus, after the third presentation cycle, participants were asked whether they were confident that they have learned the animations or whether they wanted to see the animations once more. If they responded positively, the program automatically cycled through the animations again in random order (21 participants chose this option and watched the animations a fourth time). If a participant was confident of having learned the animations, the program moved on to the next task.

The memory task was identical to that of Experiment 1. As before, correct recognition rate was equalized across conditions.

Overall correct performance was high, with correct recognition and correct rejections averaging 87% across participants. False alarms averaged 17% across participants, although across conditions, they indicated as before that the numerous and the dissimilar conditions were increasingly harder to discriminate compared to the basic condition (basic = 4%, numerous = 19%, dissimilar = 27%). From the memory data, we analyzed response latencies to YES-correct responses. Outliers above 3 standard deviations from each participant's and each condition's mean response times were removed to obtain near normal distributions (less than 5% of the analyzed data).

In the final surprise task, participants were instructed to mentally reproduce or replay the animation associated with the cue-frame in the way it exactly occurred over the original time course until it reached its original ending point. Each trial in this task started with the display of a cue-frame from those studied in the corresponding list (see Fig. 2). To minimize hand movements and time delays, participants were instructed to rest their index finger on the left-hand side button of a computer mouse throughout the task. Participants used their dominant hand to provide responses. Upon seeing the frame, they had to mentally replay the corresponding animation and click on the mouse as soon as they were done. Reproduction times were computed from the beginning of the cue frame presentation up to participants' mouse response. For each individual reproduction time, we computed a ratio score as in previous studies (Boltz, 1995). These scores represent the ratio of reproduced duration to clock duration, which indexes the extent to which the duration of the mentally reproduced event was larger or smaller than the actual event duration (a ratio of 1 indicates no difference between reproduced and actual duration). Outliers above and below 3 standard deviations from each participant's and each condition's mean ratio scores were removed from the data (39 trials out of a total of 2268 cases). Note that as in Experiment 1, analyses conducted on all scores or on scores selected based on memory accuracy revealed similar results. For the regression analyses, we simply used the reproduction times in milliseconds as the dependent variable (after outliers had been removed as done for the recognition latencies) to conduct the same regression analyses as those in Experiment 1.

2.2. Results

2.2.1. Memory judgment latencies

As the memory task was identical to that of Experiment 1, we expected to replicate these findings. As shown in Table 4, response latencies varied across condition and displayed an overall positive trend. On average, the items in the numerous condition took 77 ms longer to recognize than those in the basic condition, whereas the dissimilar condition took 109 ms longer than the basic condition. Statistical tests revealed the same pattern of results as that in Experiment 1 (a main effect of condition with log-transform latencies as dependent variable: $F_1(2, 160) = 3.04$, $p = .05$, $\eta^2 = .04$; $F_2(2, 54) = 5.11$, $p = .009$, $\eta^2 = .16$) and significant positive trends: $F_1(1, 80) = 5.24$, $p = .03$, $\eta^2 = .06$, $F_2(1, 27) = 8.85$, $p = .006$, $\eta^2 = .25$ with no significant residual variance: $F_1(1, 80) = 0.5$, $p = \text{n.s.}$,

Table 4
Response latencies and mental reproductions in Experiment 2.

Task and measure	Condition		
	Basic	Numerous	Dissimilar
Recognition – correct trials (ms)	1499 (22.21)	1576 (27.65)	1608 (28.83)
Mental reproduction (ratio scores)	1.12 (0.01)	1.14 (0.01)	1.17 (0.01)

Note: Standard errors are provided in parentheses and were computed following Loftus and Mason (1994).

$F_2(1,27) = 0$, $p = n.s.$). These results replicate the findings of Experiment 1 and therefore suggest that encoded structural properties of the animations are activated during recognition, with more event properties leading to longer response latencies.

2.2.2. Mental reproduction task

We expected that the time taken to mentally reproduce an event relative to its actual clock duration (ratio scores) would increase across the basic and dissimilar conditions as before, due to more information being stored across conditions as a function of sub-event numbers and sub-event dissimilarity. The mean ratio scores per condition are displayed in Table 4 and indeed showed that the dissimilar condition took longer to mentally reproduce than the basic condition, with the numerous condition located somewhere in-between. In milliseconds, items in the basic condition took on average 6428 ms to be mentally replayed, those in the numerous condition took 6497 ms and those in the dissimilar condition took 6613 ms. The overall pattern of statistical results resembled those reported above using ratio scores as dependent variable: a main effect of condition ($F_1(2,160) = 3.64$, $p = .03$, $\eta^2 = .04$; $F_2(2,54) = 4.16$, $p = .02$, $\eta^2 = .13$) and significant positive trends ($F_1(1,80) = 6.14$, $p = .02$, $\eta^2 = .07$; $F_2(1,27) = 6.28$, $p = .02$, $\eta^2 = .19$) with no significant residual variance ($F_1(1,80) = 0.33$, $p = n.s.$, $F_2(1,27) = 0.33$, $p = n.s.$). These results suggest that event properties encoded in memory are retrieved during mental reproduction, with reproduction times for events of the same clock duration increasing as a function of conditions.

2.2.3. Regression analyses

To evaluate the individual contribution of sub-event and similarity properties in explaining event reproduction times, we conducted by-item hierarchical multiple regressions similar to those reported in Experiment 1. We examined the proportion of variance accounted for by sub-event and similarity scores (those obtained in the stimulus pre-tests), after clock duration was taken into account in the regression model. The first step of the regression model thus included clock duration as predictor and mean reproduction times per item as the dependent variable. We found that adding sub-event scores to this regression model significantly increased the proportion of variance accounted for: R^2 increased from .66 to .75 ($F_{\text{change}}(1,81) = 29.92$, $p < .001$). In the next step, we found that the addition of similarity scores to the model also significantly increased the proportion of variance accounted for: R^2 increased from .75 to .77 ($F_{\text{change}}(1,80) = 6.13$, $p = .02$). This pattern of significance remained, regardless of the order in which predictors were entered into the model and regardless of whether means or medians from the pre-test studies were used. Table 5 provides the statistics for the full model. Thus, both the number of perceived sub-events and sub-event similarity play a role in event mental reproductions once clock duration is taken into account, suggesting that participants mentally replayed the events based on the event structure encoded in memory.

Table 5
Multiple regression model for the mental reproduction results of Experiment 2.

	Model	B	SE B	β
1.	Constant	3147.69	285.63	
	Clock duration	563.91	45.16	.81**
2.	Constant	2811.45	253.16	
	Clock duration	374.04	52.08	.54**
	Number of sub-events	314.78	57.54	.41**
3.	Constant	3602.59	402.93	
	Clock duration	361.74	50.75	.52**
	Number of sub-events	314.14	55.80	.41**
	Similarity	-152.87	61.74	-.14*

* Indicates $p \leq .02$ and ** Indicates $p < .001$.

3. General discussion

Overall, our results suggest that for a given animation, its sub-event and similarity structure modulates how we remember or reconstruct the event's temporal unfolding: more sub-events and less similar ones led to longer recognition latencies, duration attributions and mental reproductions. Moreover, the number of sub-events and their similarity accounted for orthogonal portions of variance across items, suggesting separate contributions of each event property. Although our memory-based judgments (recognition memory and duration attributions) appear different in nature from our mental replaying task, the results across all three tasks are remarkably consistent, suggesting similar influences emerging from learning. Dynamic events, like real world events, are multidimensional entities containing objects, motion and spatial information, along with cause-effect and other contingency relations. Any of these dimensions, and particularly, a change in any of them, may influence people's perception of sub-event units (Liverence & Scholl, 2012; Zacks, 2004) and their degree of similarity. Therefore, our results are important because they reveal underlying mechanisms at play when learning and reconstructing past events from memory: sub-event and similarity stimulus properties mediate the recently formed representations of events and consequently, the reconstruction of the events' unfolding from memory.

These findings sharply contrast with segmentation studies investigating timing from working memory, consistent with the claim that different tasks and paradigms target different cognitive processes (Block & Zakay, 1997; Boltz, 2005; Zakay, 1993; Zakay, Tsai, Moses, & Shahar, 1994). When monitoring duration in each trial for immediate duration estimation, more perceived stimuli lead to shorter time estimations (Block & Zakay, 1997; Liverence & Scholl, 2012; Zakay, 1993; Zakay et al., 1994). Such results have been argued to stem from attention mechanisms (more stimuli appear shorter because attention is diverted from time keeping to stimulus processing) or from working memory mechanisms (working memory updating resulting in degraded representations). In retrospect, however, as shown here, the more segments are remembered to have occurred, the longer the duration judgments, due to reliance on the encoded episodic information. This therefore suggests that event segmentation, as proposed by event segmentation theory (Kurby & Zacks, 2008), modulates encoding and retrieval in different ways as function of task and memory mechanisms involved. Segmentation organizes the flow of experience and event boundaries provide an opportunity to update working memory during ongoing processing (Radvansky & Copeland, 2006; Swallow, Zacks, & Abrams, 2009). Perhaps because of this updating mechanism during processing, episodic encoding (if required by the task) is organized according to the perceived event structure (Zacks et al., 2007). Later retrieval of these memories is then more costly as a function of the amount of episodic detail recalled (Addis & Schacter, 2008).

Our results contribute to retrospective time research by suggesting that for spontaneously structured dynamic events, participants provide duration judgments according to the encoded segmental and similarity structure. Our segmentation results resonate with early segmentation or chunking approaches to time judgments (Ornstein, 1969; Poynter, 1983, 1989) in that the number of structuring cues in the stimuli—e.g., the number of tones (Ornstein, 1969) or the number of famous names to attend to within a list of words (Zakay et al., 1994)—correlates with longer duration judgments of the stimuli. These previous results have been cast in terms of stimulus number, complexity or contextual task changes, e.g., attended names interspersed between common words. Although one could conceive our manipulations in these terms, complexity or change are vague concepts that can take many forms. Instead, we attempted to specify naturally occurring

properties of dynamic events that matter for event cognition, as perceived by naïve observers.

More importantly, our results suggest that the individual nature of the spontaneously encoded event units (e.g., color, shape, path changes) and their similarity to the contexts in which they occur play a role over and above the number of segments, not just in retrospective judgments but also in recognition memory and in the mental replay of an event's unfolding. This is important because contextual similarity (or dissimilarity) has long been recognized to modulate category formation during development (Sloutsky, 2003) and stimulus segmentation in statistical learning (Avrahami & Kareev, 1994; Gómez & Gerken, 2000). Such research suggests that similarity is fundamental in determining segment identification in the first place, the encoded structure and learning in general (Goldstone, 1994). Indeed, it is possible that an event unit is perceived as such because it differs from its immediate context and/or re-occurs in different contexts (Avrahami & Kareev, 1994), particularly when the stimuli are entirely novel as speech and objects are for a newborn. As knowledge of causal relationships in the world accumulates, event units, like those identified here, may resemble existing knowledge (e.g., a ball bouncing off a square) and thus segmentation does not only rely on bottom-up similarity but also on previous experience. Nevertheless, similarity continues to play a role in structuring unfamiliar sequences of events—e.g., one repeating schema vs. different non-repeating ones as shown here. This previous research together with the present results therefore argues for common similarity-based structuring principles shaping memory formation and the resulting memory content.

Our results also provide insights into how we build memory representations of ordinary events, even though our stimuli are relatively short compared to many ordinary events. This is because the spontaneous encoding mechanisms and structural properties revealed here are likely to operate when we first encounter novel events during development and build schemas over time in semantic memory. Patterns of co-variation in our experience, ultimately grounded in similarity between recurrent segments and their contexts (Goldstone, 1994), are already known to lead to schema extraction and category formation in cognitive development (Brady et al., 2009; Fiser & Aslin, 2005; McClelland, 2013; Orbán et al., 2008; Sloutsky, 2003). Moreover, language studies using linguistic expressions as cues to retrieve event schemas also suggest a role for the segmental and similarity structure of semantic memories. For example, people take longer to process, and attribute longer durations, to descriptions associated with less similar situation contexts in semantic memory (Coll-Florit & Gennari, 2011). Similarly, people attribute longer durations to event descriptions which have more component sub-events, such as *building a cabinet* vs. *opening a cabinet* or *painting a room* vs. *painting a house* (Burt & Kemp, 1991; Coll-Florit & Gennari, 2011; Joergensen, 2008). This suggests that event schemas (including temporal unfolding) are also based on the event structural properties of the real world. Therefore, it appears that the event properties revealed here not only hold for recently formed episodic representations but also for event schemas in semantic memory.

More generally, our approach calls for a more systematic investigation of the memory mechanisms involved in remembering both familiar and novel dynamic events and their unfolding. Surely, more research is needed to specify the possible bottom-up properties modulating encoding, and particularly, the role of similarity in learning dynamic events. Likewise, event schemas in semantic memory include information other than that emerging from a specific sequence, such as the multiple contexts in which an event can occur, which remains largely unexplored. Finally, the role of hierarchical structure—e.g., smaller units embedded into larger ones (Hanson & Hirst, 1989; Zacks et al., 2001)—in reconstructing

an event's temporal unfolding remains an open issue. Nevertheless, our work provides novel insights into how we encode and reconstruct the unfolding of dynamic events by establishing the role of encoded sub-event and similarity structures: more sub-events and less similarity between these leads to more encoded and retrieved information, and hence, to longer reconstructions of the events' temporal unfolding.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2015.06.014>.

References

- Addis, D. R., & Schacter, D. L. (2008). Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus*, 18, 1227–1237. <http://dx.doi.org/10.1002/hipo.20405>.
- Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition*, 53, 239–261. [http://dx.doi.org/10.1016/0010-0277\(94\)90050-7](http://dx.doi.org/10.1016/0010-0277(94)90050-7).
- Bellezza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 990–997. <http://dx.doi.org/10.1037/0278-7393.15.5.990>.
- Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, 4(2), 184–197. <http://dx.doi.org/10.3758/BF03209393>.
- Boltz, M. G. (1992). Temporal accent structure and the remembering of filmed narratives. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 90–105. <http://dx.doi.org/10.1037/0096-1523.18.1.90>.
- Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics*, 57(7), 1080–1096. <http://dx.doi.org/10.3758/BF03205466>.
- Boltz, M. G. (2005). Duration judgments of naturalistic events in the auditory and visual modalities. *Perception & Psychophysics*, 67(8), 1362–1375. <http://dx.doi.org/10.3758/BF03193641>.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4), 487–502. <http://dx.doi.org/10.1037/a0016797>.
- Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics*, 57(1), 105–116. <http://dx.doi.org/10.3758/BF03211853>.
- Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance*, 28(3), 600–615. <http://dx.doi.org/10.1037/0096-1523.28.3.600>.
- Burt, C. D., & Kemp, S. (1991). Retrospective duration estimation of public events. *Memory & Cognition*, 19(3), 252–262. <http://dx.doi.org/10.3758/BF03211149>.
- Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in language comprehension. *Cognitive Psychology*, 62(1), 41–79. <http://dx.doi.org/10.1016/j.cogpsych.2010.09.002>.
- Faber, M., & Gennari, S. P. (2015). Representing time in language and memory: The role of similarity structure. *Acta Psychologica*, 156, 156–161. <http://dx.doi.org/10.1016/j.actpsy.2014.10.001>.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458–467. <http://dx.doi.org/10.1037/0278-7393.28.3.458>.
- Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, 134(4), 521–537. <http://dx.doi.org/10.1037/0096-3445.134.4.521>.
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2), 125–157. [http://dx.doi.org/10.1016/0010-0277\(94\)90065-5](http://dx.doi.org/10.1016/0010-0277(94)90065-5).
- Gómez, R., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4(5), 178–186. [http://dx.doi.org/10.1016/S1364-6613\(00\)01467-4](http://dx.doi.org/10.1016/S1364-6613(00)01467-4).
- Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin*, 127(1), 22–44. <http://dx.doi.org/10.1037/0033-2909.127.1.22>.
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, 72(3), 561–582. <http://dx.doi.org/10.3758/APP.72.3.561>.
- Hanson, C., & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General*, 118(2), 136–147. <http://dx.doi.org/10.1037/0096-3445.118.2.136>.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart and Winston.
- Joergensen, G. (2008). *Visual reflections of language-mediated internal representations of time*. MSc thesis, University of York.

- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. <http://dx.doi.org/10.1016/j.tics.2007.11.004>.
- Liverence, B. M., & Scholl, B. J. (2012). Discrete events as units of perceived time. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 549–554. <http://dx.doi.org/10.1037/a0027228>.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490. <http://dx.doi.org/10.3758/BF03210951>.
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, 57(4), 335–384. <http://dx.doi.org/10.1016/j.cogpsych.2008.02.004>.
- McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, 142(4), 1190–1210. <http://dx.doi.org/10.1037/a0033812>.
- Melton, A. W. (1967). Repetition and retrieval from memory. *Science*, 158(3800), 532. <http://dx.doi.org/10.1126/science.158.3800.532-a>.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits in our capacity for processing information. *Psychological Review*, 63, 81–97. <http://dx.doi.org/10.1037/h0043158>.
- Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), 2745–2750. <http://dx.doi.org/10.1073/pnas.0708424105>.
- Ornstein, R. E. (1969). *On the experience of time*. Harmondsworth, England: Penguin.
- Poynter, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition*, 11(1), 77–82. <http://dx.doi.org/10.3758/BF03197664>.
- Poynter, W. D. (1989). Judging the duration of time intervals: A process of remembering segments of experience. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 305–332). Amsterdam: Elsevier Science Publishers.
- Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: situation models and experienced space. *Memory & Cognition*, 34(5), 1150–1156. <http://dx.doi.org/10.3758/BF03193261>.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510–532. <http://dx.doi.org/10.1037/0033-2909.114.3.510>.
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6), 246–251. [http://dx.doi.org/10.1016/S1364-6613\(03\)00109-8](http://dx.doi.org/10.1016/S1364-6613(03)00109-8).
- 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), 236–257. <http://dx.doi.org/10.1037/a0015631>.
- Waldum, E. R., & Sahakyan, L. (2012). A role for memory in prospective timing informs timing in prospective memory. *Journal of Experimental Psychology: General*, 142(3), 809–826. <http://dx.doi.org/10.1037/a0030113>.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgements: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology Learning, Memory, and Cognition*, 25(6), 1415–1434. <http://dx.doi.org/10.1037/0278-7393.25.6.1415>.
- Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 356(1413), 1363–1374. <http://dx.doi.org/10.1098/rstb.2001.0939>.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441–517. <http://dx.doi.org/10.1006/jmla.2002.2864>.
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979–1008. <http://dx.doi.org/10.1016/j.cogsci.2004.06.003>.
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, 138(2), 307–327. <http://dx.doi.org/10.1037/a0015305>.
- 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(2), 273–293. <http://dx.doi.org/10.1037/0033-2909.133.2.273>.
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130(1), 29–58. <http://dx.doi.org/10.1037/0096-3445.130.1.29>.
- Zakay, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics*, 54(5), 656–664. <http://dx.doi.org/10.3758/BF03211789>.
- Zakay, D., Tsai, Y., Moses, M., & Shahaar, I. (1994). The role of segmentation in prospective and retrospective time estimation processes. *Memory & Cognition*, 22(3), 344–351. <http://dx.doi.org/10.3758/BF03200861>.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–182. <http://dx.doi.org/10.1037/0033-2909.123.2.162>.