How language and event recall can shape memory for time

Yaqi Wang

Silvia P. Gennari

Department of Psychology

University of York

United Kingdom

Corresponding author

Silvia P Gennari

Department of Psychology

University of York

York, YO10 5DD

Email: silvia.gennari@york.ac.uk

Abstract

How do we represent the duration of past events that we have conceptualized through language? Prior research suggests that memory for duration depends on the segmental structure perceived at encoding. However, it remains unclear why duration memory displays characteristic distortions and whether language-mediated encoding can further distort duration memory. Here we examine these questions and specifically ask whether the amount of event information recalled relative to the stimulus duration explains temporal distortions. In several studies, participants first studied animated stimuli described by phrases implying either fast or slow motion (e.g., *a mule* vs *car going up a road*). They then mentally reproduced the stimuli from memory (as if replaying them in their minds) and verbally recalled them. We manipulated the amount of stimulus study and the type of recall cue (visual vs. linguistic) to assess the role of language and information recalled on the length of mental reproductions. Results indicated that the density of the information recalled (number of details recalled per second) explained temporal distortions: higher density events were lengthened and lower density events were shortened. Moreover, language additionally lengthened or shortened duration reproductions when phrases cued the task, suggesting that episodic details and verbal conceptual features were combined during recollection rather than encoding. These results suggest that the density of the details recalled and language-mediated recollection shape memory for event duration. We argue that temporal memory distortions stem from event encoding and retrieval mechanisms. Implications of these findings for theories of time, memory and language are discussed.

We remember and talk about events as unfolding over time. When recalling an event that we have recently experienced, for example, a child building a toy house, we are often able to mentally reproduce the actions or steps followed, as if replaying the events in our mind. Likewise, the language we use to talk about events reflects their temporal development. An event that we have seen, for example, can be described as *walking, strolling* or *running,* each word implying a different way in which the event unfolded. This ability to represent the unfolding of events in memory and language is a hallmark of human cognition.

Nevertheless, key aspects of this ability remain poorly understood, in particular, how event memories map onto real clock time and how this mapping is modulated by language. Indeed, we do not perceive and remember objective time in the same way we perceive and remember objects and colors, unless we pay attention to clocks. What we typically perceive, remember and talk about are events unfolding over time (Tulving, 1984). From these linguistic and memory representations, we may be able to infer temporal relations (e.g., walls are built before roofs) or details about the event unfolding (e.g., events categorized as *hurried* imply faster pace). But these internally constructed event representations are not replicas of our experiences but are rather temporally compressed, and thus do not often coincide with the real time it took these experiences to unfold.

In this article, we investigate the relationship between time, memory and language by examining how people recall and mentally reproduce or replay events that they have conceptualized through language. We specifically ask two main questions. First, we ask what determines the duration and clock accuracy of event reproductions from memory. Second, we ask how these reproductions are modulated by linguistic descriptions, thus potentially rendering reproductions more or less long or accurate. These questions are critical to understanding how the human mind represents time, how accurately it does so (e.g., when judging the duration of past events), and more generally, how it encodes and recollects events that are conceptualized through language. We thus aim to shed light on the cognitive mechanisms underpinning event memory and its relation to time and language.

**Memory for duration**

Compared to the vast literature on timing or time perception, in which stimuli are prospectively and deliberately timed (Block & Zakay, 1997; Grondin, 2010), few theories have been proposed to explain how we remember event duration in retrospect. Ornstein (1969) argued that we are often inaccurate in judging the clock duration of past events because, in the absence of clock information, we judge duration based on the amount of event information stored in memory. In other words, the more details that are encoded about an event, the longer the remembered duration of that event. Ornstein focused on an intuitive notion of stimulus complexity that would increase the amount of information stored in memory and showed that more complex stimuli are judged longer than simpler stimuli, even if their clock duration remains constant. Since Ornstein’s account, others have suggested that the number of segments or contextual changes and more generally, properties of the segmental structure may lengthen retrospective duration judgments, thereby introducing biases or distortions relative to actual duration (Avni-Babad & Ritov, 2003; Bailey & Areni, 2006; Block, 1978, 1982, Boltz, 1995, 2005; Faber & Gennari, 2015a; Jeunehomme, Folville, Stawarczyk, Van der Linden, & D’Argembeau, 2017; Poynter, 1983, 1989; Zakay, Tsal, Moses, & Shahar, 1994). For example, more segments in an interval are judged longer than fewer segments (Ornstein, 1969; Bailey & Areni, 2006; Faber & Gennari, 2015a,b). Moreover, dissimilar or unpredictable segments in an interval are judged longer than repetitive or predictable ones, because dissimilar and unpredictable segments are encoded as separate chunks in memory, whereas predictable ones are integrated into a structured whole (Avni-Babad & Ritov, 2003; Boltz, 1995; Faber & Gennari, 2015a).

However, surprisingly little research has been devoted to examining the relationship between the information actually recalled about an event and the duration attributed to it. This scarcity is in part due to the nature of the stimuli often examined: novel unfamiliar sequences are difficult to recall in detail as they become longer, even though they may be recognized. For example, participants are unlikely to exhaustively recall many unrelated words or novel movement patterns that they have seen only once. In such cases, remembered duration heavily relies on bottom-up cues to segmentation and gist-like structural characteristics, hence the reported influence of segmental structure and number of segments (Avni-Babad & Ritov, 2003; Faber & Gennari, 2015a; Zakay et al., 1994). Moreover, irrespective of stimulus familiarity, some types of duration judgments used in the literature may not be strongly related to the event information recalled (Zakay, 1993), because they are mediated by inferences and comparisons to a reference duration or to a clock unit (e.g. minutes), which themselves may not be accurately represented. Thus, whether or not recalled information plays a significant role in remembered duration may depend on the type of task and stimuli used.

The amount of event information retrieved or recalled is nevertheless likely to play a role in explaining duration memory as the segmental structure does, particularly when stimulus events are familiar. As argued by event segmentation theory and event memory models, event segmentation at encoding modulates the amount of event information recalled (Hanson & Hirst, 1989; Sargent et al., 2013; Swallow, Zacks, & Abrams, 2009; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). During the perception of familiar events, we naturally segment on-going experience based on prior event knowledge or event schemas (Zacks et al., 2007). These schemas guide predictions during encoding and shape the subsequently recalled information: The more segments are perceived at encoding, the more information is later recalled (Hanson & Hirst, 1989; Newtson & Engquist, 1976). Given that perceiving a greater number of segments in an interval leads to longer duration judgements for that interval, greater information recall should also produce longer duration judgements. The event information recalled is also arguably richer in episodic detail than the number of segments and thus it might be able to explain remembered duration better than normative measures of segmentation. Thus, although information storage and segmentation undoubtedly play a role in memory for duration, the information recalled may also prove an important determinant, at least for familiar events.

An important consequence of this recall-based view of remembered duration is that it provides a possible explanation for a currently unexplained phenomenon, namely, the tendency to overestimate short events and underestimate long ones. This phenomenon is known in the timing literature as Vierordt’s law or central tendency effects (Dyjas, Bausenhart, & Ulrich, 2012; Jazayeri & Shadlen, 2010). In memory-based judgments, this tendency has been difficult to explain because overestimation of short events and underestimation of long ones can be observed when only one stimulus event is judged by each participant (Roy & Christenfeld, 2008; Tobin, Bisson, & Grondin, 2010; Yarmey, 2000). Such findings preclude an explanation based on central tendency, as averaging stimulus durations across trials cannot take place. The recall-based account, inspired by Ornstein’s approach, would instead argue that people overestimate short duration and underestimate long ones because they remember proportionally more information per time unit for short events than long ones. In the present work, we investigate the predictions and implications of this account.

**Memory and language**

Linguistic phrases such as *building a house* have the power to cue prior event knowledge stored in semantic memory, often referred to as event concepts or schemas. Event schemas may contain information about typical actors, event structure and typical event features (Coll-Florit & Gennari, 2011; Ferretti, McRae, & Hatherell, 2001; Gennari, 2004; Gennari & Poeppel, 2003), and are thought to emerge from regularities extracted from experience over time. One enduring question in memory and language research is whether and how linguistic concepts shape the memory representation of an observed event. Studies investigating the role of language in cognition have shown that verbalizing an event or scene during visual encoding (often referred to as *thinking for speaking*) may sometimes lead to differential memory discrimination performance according to language-specific features (Feist & Cifuentes Férez, 2013; Feist & Gentner, 2007; Gentner & Loftus, 1979). These findings are consistent with interactive encoding accounts proposed for object and color categories (Lupyan, 2008, 2012): the presence of labels at encoding activates concept features that augment and distort via top-down feedback those visual aspects activated in a bottom-up fashion. This process leads to the encoding of distorted visual stimuli, resulting in failures to recognize previously seen stimuli or confusions with foils in recognition tasks (Pezdek et al., 1986, Lupyan, 2008, Feist & Gentner, 2007). It nevertheless remains unclear whether this mechanism underpins event memory more generally, as other studies examining concurrent verbal and visual event encoding do not show language effects on memory (Gennari, Sloman, Malt, & Fitch, 2002; Papafragou, Hulbert, & Trueswell, 2008; Trueswell & Papafragou, 2010).

By contrast, other accounts have argued that labels play a role during retrieval rather than encoding because they provide retrieval strategies to access episodic information (Alba & Hasher, 1983). For example, it has been shown that ambiguous drawings accompanied by labels at encoding are reproduced in distorted ways when labels cue subsequent recollection. However, the same drawings can be accurately recognized and discriminated from foils, suggesting that the initially encoded representation was not distorted by the labels (Carmichael, Hogan, & Walter, 1932; Hanawalt, 1937; Prentice, 1954; Rock & Engelstein, 1959). Pervasive influences of linguistic concepts at retrieval have also been observed even when linguistic concepts were not present at encoding (Hanawalt & Demarest, 1939). For example, some studies have shown that misleading questions posed to witnesses—in some cases, a week after encoding—can distort their numerical estimations of event duration or object speed. More specifically, the duration of an event or the speed of a moving object are reported as longer or faster, respectively, depending on the verbs used (e.g., *walk* vs *run*, *hit* vs *crash;* Burt & Popple, 1996; E. F. Loftus & Palmer, 1974). Although the nature of these *misinformation* effects remains controversial (E. F. Loftus, 2005), they suggest potential language influences at retrieval specifically when recollecting information relevant for event duration.

These observations suggest that several possibilities remain viable, particularly in the less studied domain of dynamic events and their duration. One possibility is that language shapes the encoded representations, as interactive encoding accounts suggest. However, much previous research showing language effects has used recognition memory tasks in which discrimination between highly similar items and foils is used to test memory. Such tasks could potentially cause interference between new and old items and increase uncertainty about what was actually seen, shifting the explanation onto retrieval processes or combined influences of encoding and retrieval mechanisms. Therefore, if encoded event representations are indeed augmented or distorted by the conceptual information conveyed in verbal descriptions, the influence of language should be observed in tasks that do not require discrimination between highly similar stimuli. This possibility has so far not been demonstrated for event stimuli.

Another possibility is that language influences retrieval processes by shaping the representations that are retrieved (Alba & Hasher, 1983). This shaping may take place in a way similar to that proposed for interactive encoding accounts: top-down influences from linguistic concepts, which activate a host of features stored in semantic memory, are combined with other features stored in episodic memory, giving rise to hybrid event representations. We will call this view the *interactive retrieval account*. However, it is also possible that the influence of language at retrieval is restricted to situations in which episodic memory representations are relatively weak or less reliable. Weak episodic traces would explain why there may be overreliance on language cues in some cases, as in those studies using misleading questions (Loftus and Palmer, 1974; Burt and Popple, 1996). In these cases, event information pertaining to speed and duration may not have been effectively encoded during the single event observation, or may have simply been forgotten. As a result of this weak encoding or forgetting, language may have been the most reliable cue at hand. Thus, a strong test of the interactive retrieval account should show that language modulates duration judgments in situations in which strong well-learned episodic memories are available.

**Hypotheses and overview of current studies**

The main aim of this work was to investigate the relationship between language-induced event conceptualizations, event information (or details) recalled and event reproductions. We specifically examined whether language-induced conceptualizations at encoding modulate later event reproductions or mental replays, and whether the reproduced event duration is at least partially explained by the amount of information recalled, as implied by the recall-based view and event segmentation theory. To this end, we conducted a series of studies manipulating language, frequency of stimulus exposure and retrieval cues (visual or linguistic). In particular, we varied the linguistic descriptions accompanying the visual stimuli and the number of times in which the stimuli were studied. This latter manipulation allows us to assess the role of the amount of information recalled for the same stimuli, as more stimulus study typically results in more details recalled.

In all studies, participants were first asked to study cartoon-like animations and accompanying descriptive phrases for a later memory test. The animations varied in duration and showed geometric figures moving in a familiar setting. Each animation was paired with one of two possible descriptions implying either fast or slow motion speed, for example, *a rocket being launched into the sky* vs *a Chinese lantern raising up into the sky*. Thus, the two phrases implied a shorter or longer event duration. The descriptions provided critical information to understand the animation, which would otherwise be unspecific as to the nature of the moving object. See Figure 1. After a short distraction task, participants were asked to replay the animations in their minds *exactly as they occurred in their original time course* when prompted with either an animation frame (e.g., the cue frames in Figure 1) or the corresponding description. Participants clicked the mouse at the start and finish of their mental replays. Finally, they were asked to verbally recall as many details as they could remember about each animation when prompted.



Figure 1: Examples of the experimental stimuli used in Experiments 1 to 4. Descriptive phrases are shown on the top. The small squares show a still frame near the beginning of the animation (the cue frame). The large squares represent the animations, where the arrows indicate the motion paths.

We adopted an event reproduction task because this task involves mapping the events stored in episodic memory into a mental replay that must unfold over clock time and thus captures reproduced duration independently of knowledge or inferences about clock units or familiar clock durations. This task is also similar in nature to duration reproductions and episodic simulations used in time and memory research respectively (Addis & Schacter, 2008; Grondin, 2010; Schacter & Addis, 2007). As a measure of the amount of information (or the number of details) recalled, we simply counted the number of words produced in describing the animations in detail (verbal recall task). Although this measure is admittedly crude, it yielded almost identical results to an alternative method whereby the number of details described were coded by hand. Indeed, word counts strongly correlated with the number of hand-coded details (see Supplemental Materials). Word counts also have the advantage of being objective and do not depend on arbitrary definitions of what should count as an episodic detail, thus providing a handy approximation to how much people remember about an event.

Following the recall-based account and event segmentation theory, we reasoned that if the amount of event information recalled explains duration reproductions, the number of words produced in event recall should significantly predict the reproduced duration, over and above factors already known to influence reproduced duration, such as stimulus duration and normative measures of segmentation (Faber & Gennari, 2015a, 2015b, 2017). Importantly, when the amount of exposure to the same stimuli is increased (as in Experiments 2 and 4), we expected that more event information should be learned, and thus, the reproduced duration should lengthen particularly for those events in which reproduction accuracy could be improved. In this view, the under-reproduction of long events and over-reproduction of short ones may emerge from the information recalled for each event: people may remember proportionally more information per time unit for short events, leading to lengthened duration reproductions, whereas they remember proportionally less information per time unit for long events, leading to shortened reproductions.

Moreover, we reasoned that if descriptive phrases have the power to influence how the events are encoded or retrieved, participants should additionally reproduce the event duration according to the speed implied by the phrase irrespective of retrieval cue (visual or linguistic), i.e., phrases implying faster speed should lead to shorter event reproductions than phrases implying slower speed. Specifically, as argued by the interactive encoding view, duration reproductions should be shorter or longer as a function of language in ways that are consistent with the event information recalled, thereby indicating that the event information recalled and the linguistic concept have been combined, i.e., both language and the amount of information recalled should modulate event reproductions. We test this prediction in Experiments 1 and 2. Alternatively, if language exerts an influence at retrieval, as suggested by the interactive retrieval account, evidence of combined memory and linguistic representations should be observed only when language mediates retrieval. However, linguistic labels may be relied upon only when the strength of the event information encoded is relatively weak or event details relating to motion speed are not easily accessible. Therefore, if the interactive retrieval account holds more generally, language should modulate event reproductions even when the stimulus events have been deeply encoded after repeated exposure. We test these predictions in Experiments 3 and 4.

**Experiments 1A and 1B: event reproductions after one stimulus viewing**

We began our investigation by examining the role of language and recalled information in situations in which participants were exposed to the stimulus animations only once, as it would normally be the case for events encountered in daily life, e.g., video clips watched on the web. The experiments had the general design introduced above, with a learning phase followed by the event reproduction task. In this task, participants saw the cue frame, as in Figure 1, and click the mouse to start and finish the mental replay of the animation. Experiments 1A and 1B only differed in the last task that participants performed. In experiment 1A, participants were asked to write down everything they could remember about the animations (verbal recall task), whereas in experiment 1B, they were asked to write down the phrase that accompanied the animation upon seeing the cue frame. This latter task was intended to assess whether participants indeed recalled the stimulus phrases after having seen them only once.

The interactive encoding view and the recall-based view predict that event reproductions should vary as a function of language and be explained by the information recalled over and above the influence of the stimulus duration or the number of stimulus segments. Critically, the recall-based approach predicts that the extent to which reproduced duration deviates from the actual duration—a measure typically expressed as the ratio score between the reproduced and the actual stimulus duration—should also be explained by the information recalled. In particular, more information recalled per time unit should lead to over-reproductions, i.e., reproducing the events as longer than they actually were, whereas less information remembered per time unit should lead to under-reproductions, i.e. reproducing the events as shorter than they actually were.

**Methods**

**Participants**

Based on a previously reported study most similar to the present one, namely, Experiment 2 in Burt & Popple (1996), we expected a medium effect size of *d* = .46 (Cohen, 1992). A priori power calculations with this effect size indicated that a sample of 53 participants would be sufficient to observe differences in a paired comparison (with α = 0.05 and power = 0.95). Several recent studies have also reported subtle event structure effects in duration judgments, episodic reproductions and language comprehension with 50 participants, 25 for each list in the experimental design (Coll-Florit and Gennari, 2011; Faber and Gennari, 2015a,b, 2017). We therefore aimed to recruit around 52 participants, 26 for each of our two stimulus lists (see below for the allocation of stimuli to lists).

112 English native speakers were recruited from the University of York, 58 for Experiment 1A and 58 for experiment 1B. Participants were awarded course credits or a small payment. Since the mental reproduction task is fairly unconstrained, participants could simply click through the trials without attempting to mentally replay the stimulus animations. We thus sought to identify and exclude participants who had many trials located at the extreme ends of the distribution (below or above the 2nd and 98th percentile of the distribution), because this distribution would indicate that they either clicked through the task (very short responses) or waited for a while in each trial (very long responses), and therefore did not follow instructions. If a participant had more than 30% trials falling within the extreme ends of the distribution, she/he was excluded from the data set. Six participants were thus excluded in Experiment 1A mostly due to responses below 900ms. The total number of participants for Experiment 1A was thus 52 (26 per list). For experiment 1B, four participants were excluded for comparisons of reproduced duration across experiments 1A and 1B (56 participants in Experiment 1B were entered into this comparison). Phrase recall performance in Experiment 1B, however, was assessed for all participants (58 total) because we were interested in establishing how often the phrases were recalled independently of reproductions. All studies reported here were approved by the Ethics committee of the Psychology Department (University of York).

**Stimuli**

21 cartoon-like animations were created using Adobe Flash software. Animations varied from 3s to 9s in intervals of 1s and consisted of simple geometric figures moving within a familiar setting. There were three animations for each of the seven possible stimulus duration. Each animation was paired with two descriptive phrases providing information about who or what the geometric figure was meant to represent (Figure 1). The description thus indicated what type of object the figure was, thereby implying different motion speeds (slow or fast moving entity). In seven animations, the speed information was conveyed by the verb (e.g., *children* *walking* vs *running to their classrooms*)*.* The cue frames and accompanying phrases for all animations are listed in the Supplemental Materials. Example animations can be found at https://sites.google.com/york.ac.uk/stimuli/. Each animation was also paired with a still frame near the beginning of the animation that was later used as a cue to recollect the corresponding animation (see Figure 1). Stimulus creation was guided primarily by the need to obtain animations that could plausibly represent either a slow or fast moving object with the same actual speed and similarly familiar events. An additional constraint was that the same number of items should be allocated to each possible duration.

To guarantee that the labels, the animations and the label-animation pairings only differed in the intended speed manipulation, we conducted a series of stimulus pre-tests, which are described in detail in the Supplemental Materials. These pre-tests confirmed that (a) the phrases conveyed different implied motion speed (motion speed rating); (b) the phrases fitted the animations equally well, i.e., the motion speed shown was a possible motion speed for the two alternative descriptions (phrase-animation fit rating), (c) the familiarity of the event described by the phrases was comparable across speed conditions (familiarity rating), and (d) the scale at which the scene was perceived from the viewers’ point of view, did not differ across conditions (perceived scene scale). All rating studies were conducted relative to a 1-7 scale and used a similar design to the main studies reported below. Table 1 shows the main results of these rating studies. All rating studies showed no significant difference across conditions (all p’s > .10), except for the implied speed rating (t(40)=-6.82, p<.001).

Table 1: Mean rating and standard deviations from the pre-test studies

|  |  |  |
| --- | --- | --- |
| Rating Tests | Condition | |
| Slow | Fast |
| Phrase speed | 3.17(.92) | 5.12(.92) |
| Phrase-animation fit | 6.17(.39) | 6.29(.40) |

|  |  |  |
| --- | --- | --- |
| Event familiarity | 4.83(1.42) | 5.33(1.00) |
| Perceived scene scale | 3.74(1.07) | 3.92(1.13) |

Standard deviations are given in parentheses.

Since segmentationis known to correlate with remembered duration, we also obtained the average number of segments perceived in each animation to use as a predictor in our analyses. We asked participants to indicate the smallest meaningful units they saw in each animation (fine-grained segmentation, cf. Zacks et al., 2007). We adopted an untimed segmentation task as in Faber and Gennari (2015) rather than on-line segmentation during viewing because the short length of our stimuli do not allow enough time for participants to comfortably segment the animations into fine-grained units. Moreover, our manipulation requires that participants integrate the linguistic descriptions with visual stimuli to understand the nature of the events depicted. We thus asked participants to preview the stimulus animations and their descriptions before counting the number of units in them with no time constraint. Nevertheless, segmentation during viewing strongly correlated with untimed segmentation (spearman’s rho=.84, p=<.0001) (see Supplemental Materials for details). Table 2 displays the mean number of segments across animations for a given duration, the mean number of segments per second (number of segments/stimulus duration) and the number of segments in each animation. Note that animations of the same duration could vary in the number of segments. This variability illustrates the diversity of the real-world events in which the animations were based on: Events of the same clock duration need not have the same number of segments, and long events need not contain many segments. Some long events, for example, contain smooth persistent motions with no obvious segments (e.g., fish swimming in tank), whereas some short events may contain a rapid succession of changes (e.g., cutting a banana).

Table 2: Mean number of segments for stimulus durations and animations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Stimulus duration (sec) | Mean number of segments | Mean number of segments per second | Number of segments in each animation | | |
| 3 | 2.57 | 0.86 | 2.24, | 3.34, | 2.16 |
| 4 | 3.59 | 0.90 | 4.13, | 3.18, | 3.46 |
| 5 | 5.56 | 1.11 | 3.62, | 7.13, | 5.93 |
| 6 | 4.68 | 0.78 | 4.35, | 4.00, | 5.67 |
| 7 | 4.01 | 0.57 | 2.63, | 6.58, | 2.92 |
| 8 | 4.95 | 0.62 | 5.02, | 4.12, | 5.70 |
| 9 | 4.11 | 0.46 | 6.19, | 2.44, | 3.82 |

**Design and procedure**

21 animations were paired with two phrases each. Each phrase-animation pair was assigned to a different stimulus list. Each list contained 11 or 10 items in the Slow and Fast language conditions. Each participant only saw each animation in one language condition but saw all animations across conditions. The experiment consisted of two phases: a learning phase and the test phase.

**Learning phase**. Participants were instructed to study the animations and phrases for a subsequent memory test. They were told that the phrases provided a clue to understanding what was going on in the animation. They were also told that the animations were cartoon-like representations of the events described and thus the objects and scenes were not meant to have photographic fidelity. After instructions, participants saw all 21 animations in a list once in a random order. In each learning trial, participants first saw a phrase with a cue-frame underneath and then pressed a key to watch the corresponding animation. See Figure 2 for details of the trial structure. No mention of time was made at this stage.



Figure 2: Schematic representations of the learning and replay tasks in Experiment 1.

**Test phase**. After a distraction task consisting of math calculations (which lasted approximately 10 minutes), participants completed a mental replay task. In this task, participants were told to mentally replay the cued animation exactly as it occurred in its original time course. Instructions provided examples of how the task would proceed. In each trial, participants first saw a cue-frame to remind them of the animation that they were about to mentally replay. At the bottom-right corner of the screen, there was a phrase prompting the start of the replay and a mouse symbol (See Figure 2). When they were ready, they clicked the mouse to start mentally replaying the animation as veridically as they could. During the mental replay, the cue frame remained on the screen together with a mouse symbol and a prompt to click again when they were done mentally replaying the animation. The duration between the two clicks was considered as the reproduced duration for each animation. The cue-frame presentation order was random. Participants had to keep their dominant hand on the mouse throughout the task for more accurate timing.

Finally, participants performed one of two tasks. In Experiment 1A, participants were asked to type everything they could remember about the animations (verbal recall task). Instructions indicated that they should write as many details as they could remember about what happened in the animations as well as physical characteristics (color, shapes, etc.). Each trial started with the presentation of the cue-frames in random order and participants wrote their memories in a textbox that appeared underneath the frame. We used this task to obtain a measure of the amount of information recalled by counting the number of words used (see Supplemental Materials for further discussion of this measure). Although verbalizing memories could emphasize a particular way of packaging information, verbal recall has been widely used in memory research and is probably the only type of measure that can approximate the amount of information recalled. In Experiment 1B, participants were asked to write down the phrase that accompanied the animation (phrase recall task). Thus, for Experiment 1B, we do not have the number of words recalled as a predictor.

**Data treatment and statistical analyses**

Two main dependent measures were used, the reproduced duration (RD) in milliseconds and the ratio between reproduced duration and actual clock duration, which represents the extent to which reproduced duration deviates from the actual stimulus duration. We will call this measure *deviation index* (DI). A deviation index larger than 1 indicates a longer reproduced duration than the actual duration, whereas an index smaller than 1 represents a shorter reproduction relative to actual duration. To minimize the influence of outliers in our analyses, word counts from the recall task were log-transformed and reproduced durations falling below the 2nd percentile or above the 98th percentile were removed from the data set. These meant that reproductions shorter than 1000ms or larger than 13700ms were excluded. We also excluded trials for which participants did not correctly recall the events of the animation, as in these cases, there could be no relation between recalled information and reproduced duration. Verbal recall for a given trial was considered accurate if at least one correct piece of information about the animation was provided, irrespective of the level of detail provided (see attached data files). Together these exclusions comprised less than 4.5% of the data set.

Hypotheses testing in this and all subsequent experiments used linear mixed-effects models carried out in the *lme4* package of *R* (R core team, 2017)(Bates, Maechler, Bolker, & Walker, 2015). The models included the maximal random effects structure when convergence obtained (i.e., random intercepts for subjects and items, by-subject random slopes for stimulus duration, and by-item and by-subjects random slopes for the language condition) (Barr, Levy, Scheepers, & Tily, 2013). Data and the model comparisons performed can be found at https://osf.io/8ub3q/?view\_only=130d85611e574e7c9fb77a4a47f5a7c7. Effects of interest were assessed by likelihood ratio tests comparing the full model with the effect of interest to a model without this effect and maximal random effects structure. For example, to test for the effect of recalled information over and above stimulus duration and the average number of segments, we added the log-transformed number of words recalled to a model containing stimulus duration, language condition and the mean number of segments in each animation. Models explaining deviations from the stimulus duration were similarly constructed but the fixed factors were the mean number of segments per seconds (number of segments/stimulus duration) and the number of words per seconds (log-number of words/stimulus duration). Stimulus duration need not be a fixed factor in these latter models because stimulus duration is used to compute the deviation index, although it was included in the random effects structure. Since the main purpose of Experiment 1B was to assess stimulus phrase recall and no difference was observed across Experiments 1A and 1B in reproduced duration (t(105) =.35, p > .05), we focused our modeling results in Experiment 1A.

**Results**

**Reproduced Duration**

Mixed-effects modeling assessing the effect of language on reproduced duration revealed no significant effect of language (χ2(1)=.09, p=0.75), suggesting that the way in which participants encoded the events did not influence their mental reproductions. Since the deviation index is computed from the reproduced duration, there was no language effect for this dependent variable either (χ2(1)=.30, p=0.58). Figure 3A shows the range of reproduced durations as a function of stimulus duration in a Tukey’s style boxplot (the boxes’ height represents the first and third quartile of the distribution, the middle line is the median and whiskers indicate the largest and lowest values equaling ±1.5 × the interquartile range). Figure 3B shows the deviation index (error bars were adjusted as suggested by (G. R. Loftus & Masson, 1994) for repeated designs). As can be observed in Figure 3A, duration reproductions show a linear increase plateauing for animations longer than six seconds. In the majority of cases, these reproductions were not accurate. As shown in Figure 3B, shorter animations were reproduced as longer than they were, and longer animations were reproduced as shorter, thus showing reproduction biases resembling those previously reported. The final model with the main effects of interest is reported in Table 3.

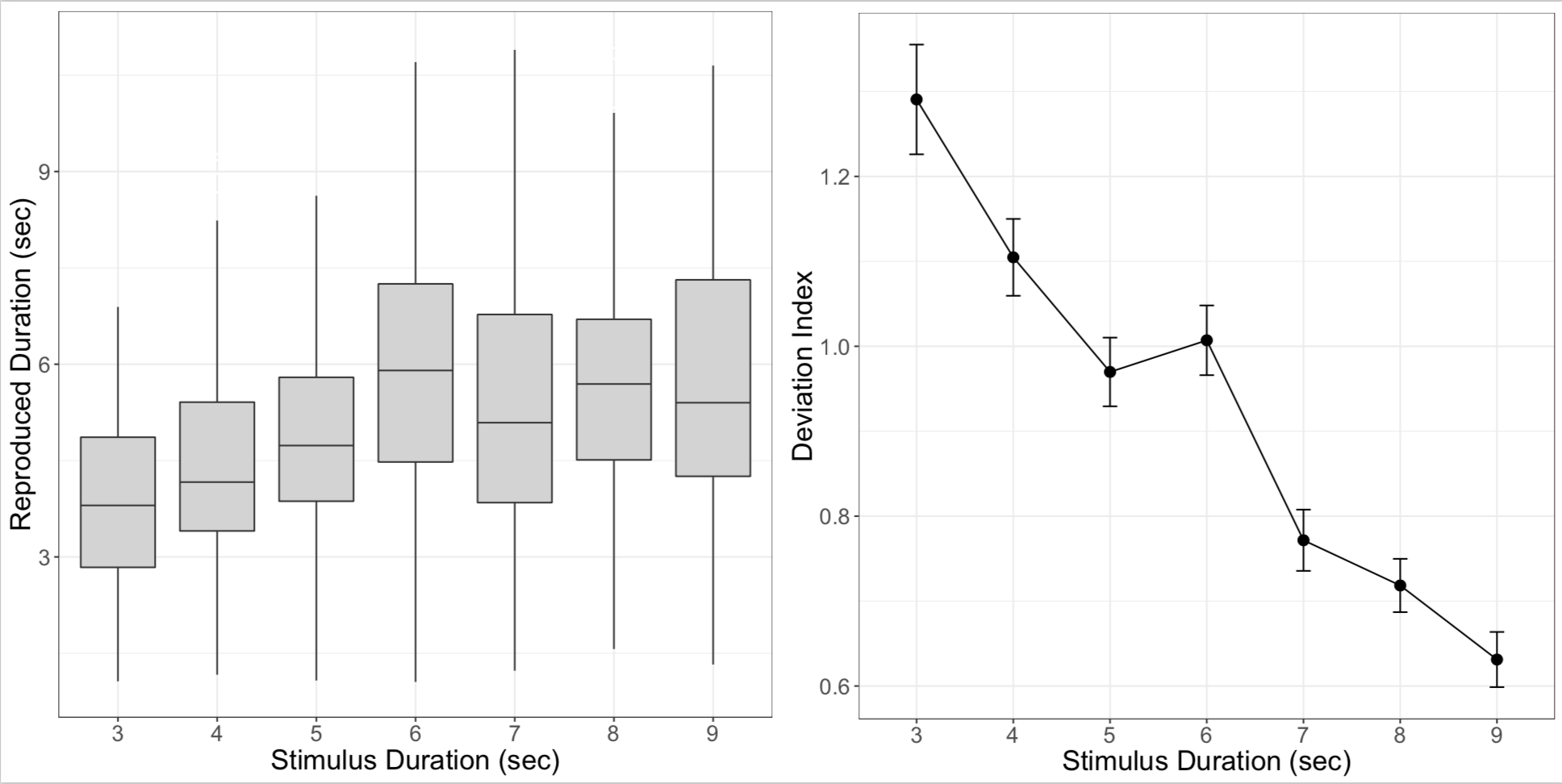


Figure 3: Results of Experiment 1. Panel A shows the medians and interquartile ranges of the reproduced durations as a function of the stimulus duration. Panel B shows the deviation index (reproduced duration/stimulus duration) as a function of the stimulus duration. Error bars indicate standard errors calculated according to Loftus and Masson (1994).

Further model comparisons assessing the role of recalled information (log-transformed number of words recalled) on reproduced duration over and above stimulus duration and number of segments revealed that the information recalled made a significant contribution to the model (χ2(1)=5.75, p=0.02): the more information recalled, the longer the reproduced duration (Table 3). Importantly, the amount of information recalled per time unit (log(words)/stimulus duration) modulated the extent to which reproduced duration deviated from the stimulus duration (deviation index), over and above the number of segments per second (χ2(1)= 9.99, p=0.002): as the number of words per second increased, so did the deviation index, indicating over-reproductions. This result suggests that in addition to segmental structure, the extent to which animations are shortened or lengthened relative to stimulus duration is explained by the information recalled per time unit—a measure of information density. Exploration of possible interactions between the number of segments, the word counts and the stimulus duration revealed no significant interaction. Figure 4A shows the relationship between the mean reproduced duration and the log-transformed number of words for each of the 21 stimulus animations. Figure 4B shows the relationship between the deviation index and the number of words recalled per seconds in the animation.

Table 3: Modeling results for Experiment 1A

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent  Variable | Fixed effects | Estimated coefficient | Standard error | t-value |
| Reproduced | (Intercept) | 2247.10 | 440.66 | 5.10 |
| Duration (ms) | Stimulus duration | 261.62 | 51.32 | **5.10** |
|  | Language Condition | 37.83 | 88.25 | 0.43 |
|  | Stimulus segments | 141.57 | 65.15 | **2.17** |
|  | Log(words) | 562.67 | 230.88 | **2.44** |
| Deviation | (Intercept) | 0.45 | 0.09 | 4.81 |
| Index | Language Condition | 0.01 | 0.02 | 0.84 |
|  | Segments per second | 0.36 | 0.11 | **3.33** |
|  | Log(words) per second | 0.74 | 0.20 | **3.68** |

Note: Bold values indicate significant fixed effects (p < .05).

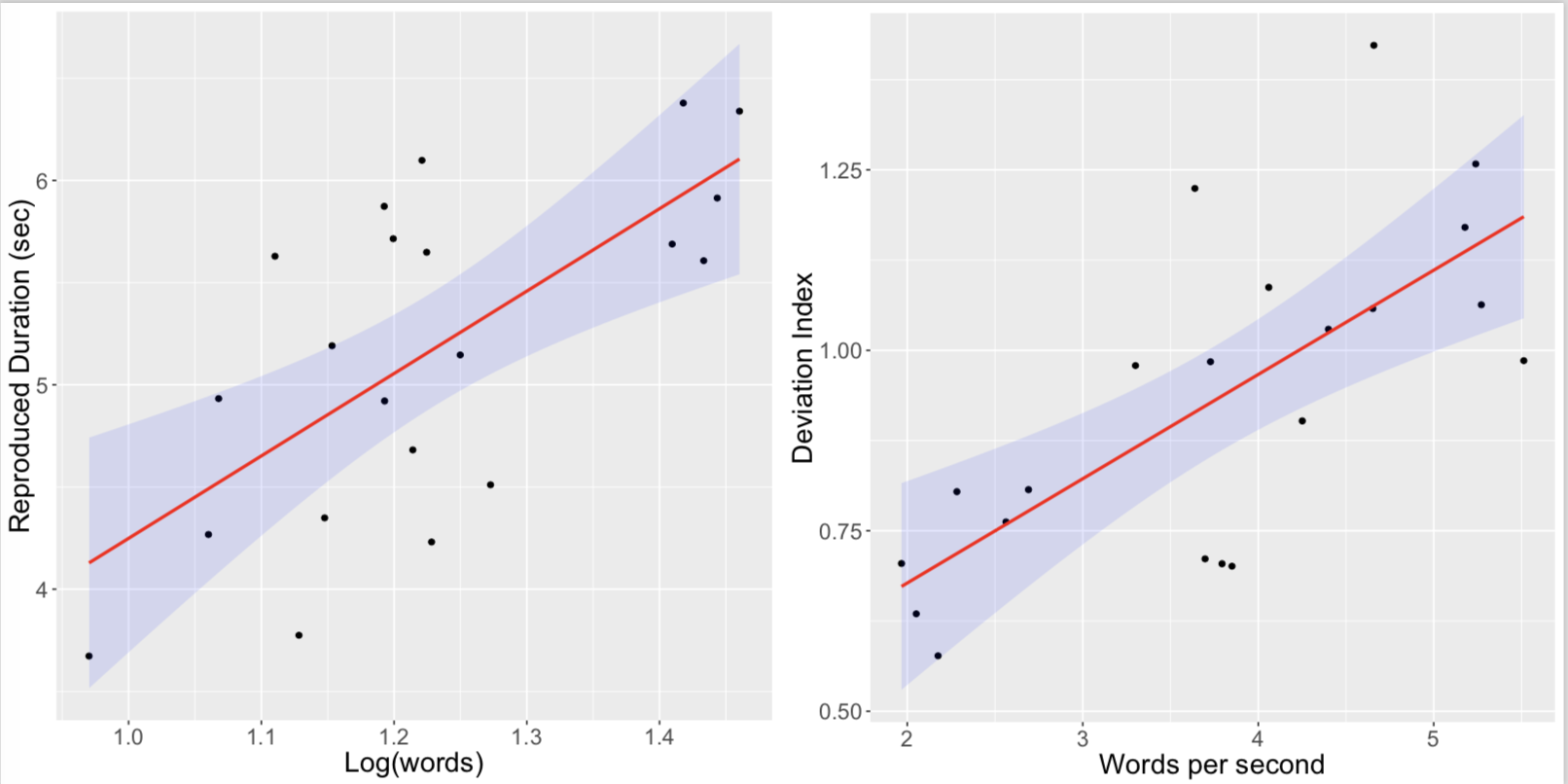
****

Figure 4: Results of Experiment 1. Panel A shows the relationship between the mean reproduced duration and the mean log-transformed number of words recalled for each of the stimulus animations. Panel B shows the relationship between the mean deviation index (reproduced duration/stimulus duration) and the mean number of words per second (number of words recalled/stimulus duration) for each stimulus animation.

**Event and phrase recall accuracy**

Participants in Experiment 1A recalled appropriate information about the animations with high accuracy. On average, they provided correct information about the animations in 95% of cases (range: 71-100%). The mean accuracy was the same across language conditions. Although participants were not required to use the stimulus phrases in their verbal recall, we inspected their descriptions to assess whether they recalled the linguistic conceptualization of the events. Manual coding of the data indicated whether participants named the moving objects with the nouns of the stimulus phrases, e.g., *rocket vs. lantern*. This coding revealed that they did so on average 72% of the time, with large variability across participants (range: 10-100%). In the remaining cases, they referred to the objects as geometric figures (e.g., *a yellow rectangle appears at the bottom of the rectangle and they both move up the screen* for the left animation in Figure 1).

To evaluate phrase recall in Experiment 1B, we automatically coded the phrases that participants reported. A program searched for the critical speed words in the phrases that would distinguish one condition from the other (e.g., *rocket vs. lantern, ambulance* vs. *bus* or *walking vs. running*). This coding was thus more stringent than that used for event recall above because the specific nouns and verbs conveying the critical speed information was searched for, rather than the nouns alone. Alternative synonyms or near-synonyms were added to the search list and thus considered correct (e.g., *runners* and *athletes*). On average, participants recalled the critical words of the phrases 85% of the time (range: 51-100%). Taken together, these results suggest that participants generally recalled the critical information in the stimulus phrases, but performance varied greatly across participants.

**Discussion**

The results of these experiments indicated no influence of language on event reproductions, even though statistical power was similar to that of previous related studies. Whether the language implied a fast or slow motion for an animation, which entails conceptualizing the main object as a fast- or slow-moving entity, had no effect on how the events were mentally reproduced, against the expectation of the interactive encoding account. This result is somewhat surprising because duration reproductions were often inaccurate, and thus participants did not encode or retrieve the true motion speed or duration. On the contrary, participants appear to have been uncertain about the pace of the animation, as suggested by the great variability of scores across individuals and animations (see Figure 3A). One possible explanation for the absence of a language effect consistent with poor reproduction accuracy is that participants did not encode language-based speed representations strongly enough to be able to retrieve them later. Although the cue frame can bring to mind the associated phrase on average 85% of the time, as suggested by the phrase recall tasks (Experiment 1B), this performance does not reflect the nature of the episodic details encoded. In this view, participants were able to provide details in the recall task or retrieve the phrases because they recalled the gist of the animations or gist-like features, but the language-induced episodic details that must be retrieved to lead to an observable language effect in duration reproductions were not really accessible. We address this possibility in Experiment 2.

The most important finding of the present studies was that the event information recalled modulated the reproduced duration. As the number of words recalled increased, so did the reproduced duration, over and above the influence of stimulus duration and the number of segments. This finding suggests that the information and the details recalled underpin event reproductions. Moreover, deviation scores (Figure 3B) indicated that shorter animations were lengthened and longer animations were shortened, a common bias in duration judgments (see Introduction). Critically, the density of the details recalled (the number of words recalled per seconds in an animation) was able to explain the extent to which reproduced duration deviated from the stimulus duration, in addition to the segmental density. This result suggests that shorter animations were reproduced as longer because more details were proportionally recalled for them compared to longer animations, thus providing a possible explanation for the biases observed. However, it remains an open question whether varying the amount of information recalled for the same stimuli would also change the reproduced duration, a stronger test for the role of information recalled on reproduced duration. We test this possibility in Experiment 2.

**Experiment 2: event reproductions after several stimulus viewings**

One possible explanation for the absence of language effect in Experiment 1 was that the phrases did not influence the episodic details retrieved because participants did not encode the animations sufficiently well, and mostly recalled gist-like aspects, rather than speed-related episodic details. To address this possibility, Experiment 2 provided participants with more opportunities to learn the animations and their labels by increasing the exposure to the stimulus set. By deeply encoding the animations according to the language, linguistically conceptualized representations might be retrieved during event reproductions. If language can modulate event reproductions under these circumstances, we would expect shorter reproductions for event labels implying fast motions compared to those implying slow motions.

Moreover, the recall-based approach to reproduced duration predicts that by studying the animations in more detail, participants should also be able to learn more about them, and thus the amount of information recalled should increase in this experiment compared to that of Experiment 1. Critically, if more information is recalled for the stimuli, this approach predicts that reproduction accuracy should also increase in this study compared to those of Experiment 1. An increase in reproduction accuracy means longer reproduced durations for longer stimuli in Experiment 2 compared to Experiment 1, but shorter or similarly long reproduced duration for shorter stimuli. Indeed, recall that participants in Experiment 1 over-reproduced the duration of the short animations, and it is unlikely that these over-reproductions would become even longer in Experiment 2 because learning should improve reproduction accuracy, rather than making reproductions more deviant. Repeated stimulus observations in varying orders surely allow for fine-grained stimulus segmentation and implicit comparisons across animations in multiple event dimensions (path travelled, objects, etc.), thus improving sensitivity to the differences between animations. Moreover, more episodic details were missed for longer animations in Experiment 1, as shown by the fewer words per second recalled (Figure 4B), and therefore, there is more room for memory improvements in these cases. The effect of exposure should thus not be the same across all animation durations, i.e., there should be an interaction between exposure and stimulus duration, with larger accuracy improvements for longer animations. Nevertheless, as in Experiment 1, we expect an overall relationship between the amount of information recalled and the reproduced duration over and above stimulus duration, because relative to the stimulus set, more information should be recalled for longer animations compared to shorter animations. Similarly, deviations from the stimulus duration (deviation index) should be predicted by the density of the details recalled (number of words per second), as found in Experiment 1.

**Methods**

**Participants**

57 English native speakers who did not participate in the previous experiment were recruited from the University of York for course credit or payment. One participant was excluded, because he/she did not complete the recall task. Four other participants were excluded from data analysis according to the exclusion criterion of Experiment 1. In total, there were 52 participants (26 participants in each stimulus list).

**Design, procedure and data analyses**

This study used the same stimuli, design and procedures as in Experiment 1, except for the learning phase. In the present learning phase, participants were shown the animation and phrases three times. Once participants had seen all animations in random order, they were told that they would watch the animations again so they could learn them in more depth. The program then cycled twice through the stimuli in random order, with a screen midway indicating that they would watch the animations once more. After the learning phase, participants performed a distraction task (math calculations), the mental replay task and the verbal recall task as before. Data treatment was as before: inaccurate responses and values falling below or above the 2nd and 98th percentile of the overall distribution were excluded. Model comparisons and the models’ random effects structures were as described for Experiment 1, with the addition of Exposure as a fixed factor (and by-items random slope if convergence allowed) when comparing across Experiments 1 and 2. Specifically, we tested for an interaction between stimulus duration and exposure, once stimulus duration and the main Exposure effect were accounted for in the base model.

**Results**

**Reproduced duration**

Model comparisons assessing the effect of language indicated that there was no language effect in reproduced durations (χ2(1)= 0.08, p=0.93) or deviation indices (χ2(1)= 0.07, p=0.78). This result thus replicates the findings of Experiment 1, despite the fact that exposure to the stimuli should have increased the strength of the linguistically conceptualized representations. As in Experiment 1, the number of words produced in the recall task was a significant predictor of reproduced duration over and above stimulus duration and the mean number of segments in the animations (χ2(1)= 5.38, p=0.02), indicating that as the word count increases, so does the reproduced duration. Similarly, the number of words per second was a significant predictor of the deviation index, over and above the number of segments per second (χ2(1)= 14.72, p=0.0001). Table 4 shows the models’ summaries. Further exploratory model comparisons indicated no significant interactions between stimulus duration, number of segments and word counts. These results thus replicate those of Experiment 1.

Table 4: Models’ summaries for the results of Experiment 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent  Variable | Fixed effects | Estimated coefficient | Standard error | t-value |
| Reproduced | (Intercept) | 1201.65 | 354.08 | 3.39 |
| Duration (ms) | Stimulus duration | 456.08 | 44.97 | **10.14** |
|  | Language Condition | -2.08 | 79.95 | -0.03 |
|  | Stimulus segments | 179.76 | 43.41 | **4.14** |
|  | Log(words) | 553.44 | 232.34 | **2.38** |
| Deviation | (Intercept) | 0.53 | 0.06 | 8.57 |
| Index | Language Condition | 0.00 | 0.01 | 0.31 |
|  | Stimulus segments per second | 0.25 | 0.06 | **4.02** |
|  | Log(words) per second | 0.88 | 0.17 | **5.23** |

Note: Bold values indicate significant fixed effects (p < .05).

**Comparisons across Experiments 1 and 2: Exposure effect**

**Event Recall Accuracy.** Verbal recall was generally very accurate in Experiment 2 (mean= 99%, range: 90=100%), suggesting that participants learned the animations very well and were able to provide information about what happened in them. Recall accuracy was better in Experiment 2 than in Experiment 1 (99% vs. 95%, Wilcoxon test: Z=-5.01, p<.001). Manual coding of the nouns used in the verbal recall task indicated that they used the stimulus words 65% of the time (range: 0-100%). This percentage was comparable to that reported for Experiment 1, as there was no difference across experiments (Wilcoxon test: Z=-1.00, p >.05).

**Number of words used in recall.** Given the increased amount of stimulus exposure, we expected that participants in Experiment 2 would produce more words in their verbal recall than participants in Experiment 1 (main effect of exposure). Using log(word count) as dependent variable (the stimulus duration was treated as fixed factor and random slope, and Experiment as random slope), statistical comparisons revealed a main effect of exposure (χ2(1)= 8.23, p=0.004), indicating that there were more words used in Experiment 2 than Experiment 1, with a numerical trend towards larger differences for longer animations (see Table 5). This result confirms that participants indeed learned more about the animations in the present experiment.

Table 5: Mean number of words produced in the recall task for Experiments 1 and 2 as a function of stimulus duration

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Stimulus | Experiment 1 | | | Experiment 2 | |
| Duration (sec) | Mean (SD) | | | Mean (SD) | |
| 3 | 13.37 | (8.28) | 16.55 | | (7.82) |
| 4 | 18.55 | (10.77) | 20.70 | | (10.05) |
| 5 | 20.64 | (11.85) | 24.85 | | (11.11) |
| 6 | 28.14 | (17.52) | 35.68 | | (16.06) |
| 7 | 16.15 | (9.48) | 19.56 | | (9.86) |
| 8 | 27.02 | (15.12) | 34.03 | | (14.35) |
| 9 | 23.92 | (16.90) | 30.58 | | (18.02) |

Note: Parentheses indicate standard deviations

**Reproduced duration**. To test whether duration reproductions would increase in Experiment 2 compared to Experiment 1 for the longer animations in the set, we examined whether there was an interaction between stimulus duration and exposure in explaining reproduced duration. Model comparisons indicated a significant interaction for reproduced durations (χ2(1)= 16.99, p<0.001) and deviation indices (χ2(1)= 12.62, p<0.002). The models are summarized in Table 6. Figure 5 shows the reproduced duration and the deviation index as a function of the stimulus duration in Experiment 1 and 2. These results indicate that duration reproductions become longer for longer animations and slightly shorter for shorter animations, whereas deviations become more accurate, i.e., closer to 1, as function of exposure. The difference across experiments was larger for longer animations, although there was also a tendency towards greater accuracy for shorter durations, i.e., smaller over-reproductions. Nevertheless, the tendency to over-reproduce shorter durations and under-reproduce longer durations was observed in both experiments, albeit to a different degree.

Table 6: Model summaries for comparisons between Experiments 1 and 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent  Variable | Fixed effects | Estimated coefficient | Standard error | t-value |
| Reproduced | (Intercept) | 3798 | 380 | 9.99 |
| Duration (ms) | Stimulus duration | **204** | **69** | **2.95** |
|  | Exposure | **-481** | **116** | **-4.14** |
|  | Exposure × stimulus duration | **103** | **24** | **4.28** |
| Deviation | (Intercept) | 1.64 | 0.09 | 17.55 |
| Index | Stimulus duration | **-0.12** | **0.01** | **-10.40** |
|  | Exposure | -0.07 | 0.03 | -2.03 |
|  | Exposure × stimulus duration | **0.014** | **0.004** | **3.47** |

Note: Bold values indicate significant fixed effects (p < .05).

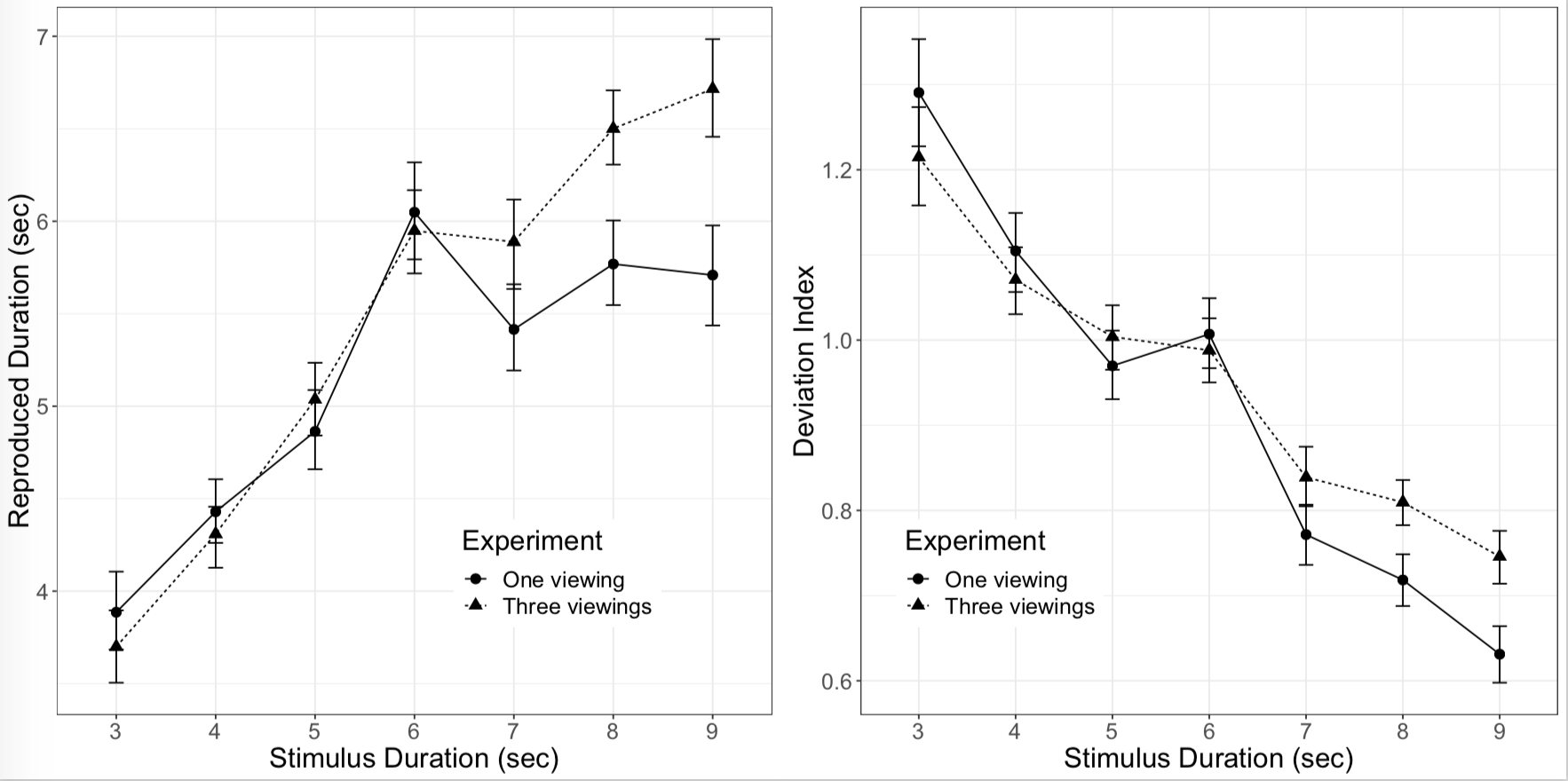
****

Figure 5: Results of Experiments 1 and 2. The panels show the reproduced duration (in A) and the deviation index (in B) as a function of the stimulus duration. Error bars indicate standard errors calculated as in Loftus and Masson (1994) for each group.

**Discussion**

The results of Experiment 2 were similar to those of Experiment 1 in that there was no effect of language. Although Experiment 2 was intended to strengthen the linguistically conceptualized encoding, the animations’ reproductions were not faster or slower as a function of language. This finding was again surprising because duration reproductions deviate from the stimulus duration in many cases so that participants do not seem to have encoded and retrieved the true animation pace. This result, therefore, is inconsistent with an interactive encoding account. As suggested by the recall-based approach, event memory was the main information source for event reproduction, and linguistic information did not shape event memories. This finding raises the question of whether language may play a role in reproduced duration when the phrases are used as cues to recall the animations, as suggested by the interactive retrieval account discussed in the Introduction. In such cases, visual information must be retrieved through language, and therefore, there is a clear opportunity to combine conceptual and episodic information. We address this possibility in the next studies.

The results of Experiment 2 were also similar to those of Experiment 1 in that the number of words produced in event recall and the density of the information recalled were significant predictors of reproduced duration and deviation index respectively, confirming the predictions of the recall-based approach. Importantly, more accurate reproductions were obtained in Experiment 2 compared to Experiment 1, with a larger effect observed for longer animations in the stimulus set. This result is consistent with the recall-based view and suggests that the number of details recalled underpins duration reproductions. Moreover, the results of Experiments 1 and 2 provide support for the hypothesis that the tendency to lengthen short durations and shorten long ones stems from the density of the details recalled. Since more information (words per seconds) is remembered per unit of time for short animation and less information is remembered for longer ones, this results in over-reproductions for short animations and under-reproductions for longer ones.

One intriguing aspect of the present results is that the reproductions of some animations such as those lasting five and six seconds tended to be on average accurate, despite great variability in individual scores, as shown in Figure 3A. This tendency might have occurred because the number of segments in these animations (between 4 and 6) and their episodic details happened to align better with the stimulus duration compared to other animations. As noted by theories of event perception and memory (Radvansky & Zacks, 2011, 2017), participants did not encode mental replicas of the animations, but rather, they encoded them in terms of changes, distance traveled, motion path, etc. Given that participants aimed to be accurate in their episodic reproductions, systematically varying the clock duration of the stimulus set is bound to yield some accurate reproductions, namely those in which constraints on information processing (e.g., how much is stored and retrieved) align with the actual stimulus duration. We will come back to this issue in the General Discussion.

**Experiment 3: event reproductions cued by language**

This study investigates whether language modulates event reproductions from memory when linguistic phrases are used as retrieval cues. Previous studies using misleading questions at retrieval have shown influences of language on numerical estimates of duration and speed when language had not been present at encoding (Loftus and Palmer, 1974; Burt and Popple, 1996). In some of these studies, the relevant questions were interspersed with requests to verbally describe the events. Descriptions of events and objects that are no longer present are known to impair subsequent memory performance. This phenomenon is often referred to as verbal re-coding or overshadowing, because previously visual representations are re-encoded in often novel conceptual terms (Chin & Schooler, 2008; Schooler & Engstler-Schooler, 1990). In contrast, the present study examines the use of phrases as recall cues after an associative link is established in memory between a phrase and an animation during learning, and therefore, post-encoding verbalizations do not play a role.

Experiment 3 asked participants to study the animations and associated phrases once, as in Experiment 1. After learning, participants were asked to mentally reproduce the animations and provide event details when prompted by the phrases. The use of phrases as cues to retrieve the animation guarantees that the conceptual information conveyed by language will be activated right before the event reproduction. Will participants be able to access episodic information independently of the phrases? If so, duration reproductions should show a similar profile as those of Experiment 1 and be unaffected by language. Alternatively, as argued by the interactive retrieval account, the conceptual and retrieved memory representations may be combined, leading to longer or shorter reproductions as a function of language but in ways that are consistent with the amount of information recalled. This possibility, therefore, predicts that there should be an effect of language as well as a relationship between reproduced durations and recalled information similar to those shown above. Likewise, the average duration reproduction independently of language in this experiment should not differ significantly from that of Experiment 1, because event memory should still drive reproductions, despite language influences.

**Methods**

**Participants**

55 English native speakers who did not participate in previous studies were recruited from the University of York for course credit or payment. In total, 3 participants were excluded from the data analysis because more than 30% of their data were outliers or forgotten animations. In total, there were 52 participants, 26 in each list.

**Stimulus, design and procedure**

The same stimulus animations, design and procedures used in experiment 1 were adopted for this study with one alteration: instead of using the cue frames as the cue to recall or mentally reproduce the animation, the corresponding phrase was used. The trial structure was also identical to that illustrated in Figure 2, except that a phrase replaced the visual cue. Similarly, the verbal recall task used the phrase rather than the cue-frame as the recall cue. Participants were presented with a phrase and a textbox underneath to enter their responses. Note that in this task, the cue phrase contains the gist of the animation so that participants are constrained to provide other details about the animations not already mentioned in the phrase. In particular, visual details that were previously present in the cue frame (colors, shapes and setting) tended to be described.

Data treatment and analyses proceeded as described for Experiment 1. Models included that maximal random effects structure allowed by convergence (by-subject random slope for the language condition and stimulus duration, and by-item random slope for the language condition). To compare across experiments, we created a model containing stimulus duration and added the experiment (Experiment 1 vs Experiment 3) as fixed factor and by-items random slope. Overall, participants in this study were accurate in providing details about the animations beyond the information provided by the phrase, e.g., the color of the objects (mean accuracy: 98%, range: 81%-100%). If no correct detail was provided, recall was considered inaccurate.

**Results**

**Reproduced Duration**

Model comparisons revealed significant effects of Language for the reproduced duration (χ2(1)= 8.05, p=0.005) and the deviation index (χ2(1)= 8.21, p=0.004), indicating that reproductions were longer for phrases implying slow motion compared to those implying fast motion. This effect was similar across clock durations, as there was no interaction between language and stimulus clock duration. See Figure 6A. On average across animations, reproduced durations were 279ms longer in the Slow condition than the Fast condition (mean Slow condition = 5478ms, mean deviation index = .99; mean Fast condition = 5199ms, mean deviation index = .93). The distribution of deviation scores was similar to that of Experiment 1 (see Figure 6B) but language slightly shifted them up or down. That is, over-reproductions or under-reproductions were slightly lengthened or shortened according to language. These results suggest that when language is used to retrieve event memories, conceptual information shapes how events are mentally reproduced to have unfolded.

Models assessing whether the event information recalled still bears a relationship with reproduced duration and deviation indices, despite the change in cue-recall, revealed a significant contribution for reproduced duration (χ2(1)= 15.79, p<0.001) and deviation indices (χ2(1)= 9.52, p=0.002), indicating that as the number of words and the number of words per second (information density) increase, so do the reproduced duration and the deviation index. See Table 7 for model summaries. There were no interactions between language, number of segments or recalled information in these models. We also examined whether duration reproductions were similar across Experiment 1 and 3 when averaging across the language conditions (effect of Cue Type), since the average reproduced duration per stimulus duration should not drastically differ across experiments, if similar event memories underpin all event reproductions. There was indeed no significant difference in comparing reproduced durations or deviation scores across experiment 1A and Experiment 3 when collapsing across phrase conditions (duration: χ2(1)= 0.11, p>0.05; ratio: χ2(1)= 0.88, p > .05). As shown in Figure 6B, the overall pattern of results (averaging across language conditions) was very similar to that reported in Experiment 1, with a tendency to over-reproduce short durations and under-reproduce long ones. These results suggest that despite language influences, the information recalled played a similar role in the reproduced duration as in previous experiments.

Table 7: Model summaries for the results of Experiment 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent  Variable | Fixed effects | Estimated coefficient | Standard error | t-value |
| Reproduced | (Intercept) | 2096.53 | 503.38 | 4.17 |
| Duration (ms) | Animation duration | 289.98 | 56.27 | **5.15** |
|  | Language condition | -266.31 | 93.41 | **-2.85** |
|  | Animation segments | 177.20 | 73.58 | **2.41** |
|  | Log(words) | 885.21 | 220.48 | **4.02** |
| Ratio score | (Intercept) | 0.56 | 0.09 | 6.03 |
|  | Language condition | -0.05 | 0.02 | **-2.97** |
|  | Animation segments per second | 0.35 | 0.11 | **3.28** |
|  | Log(words) per second | 0.69 | 0.19 | **3.54** |

Note: Bold values indicate significant fixed effects (p < .05).

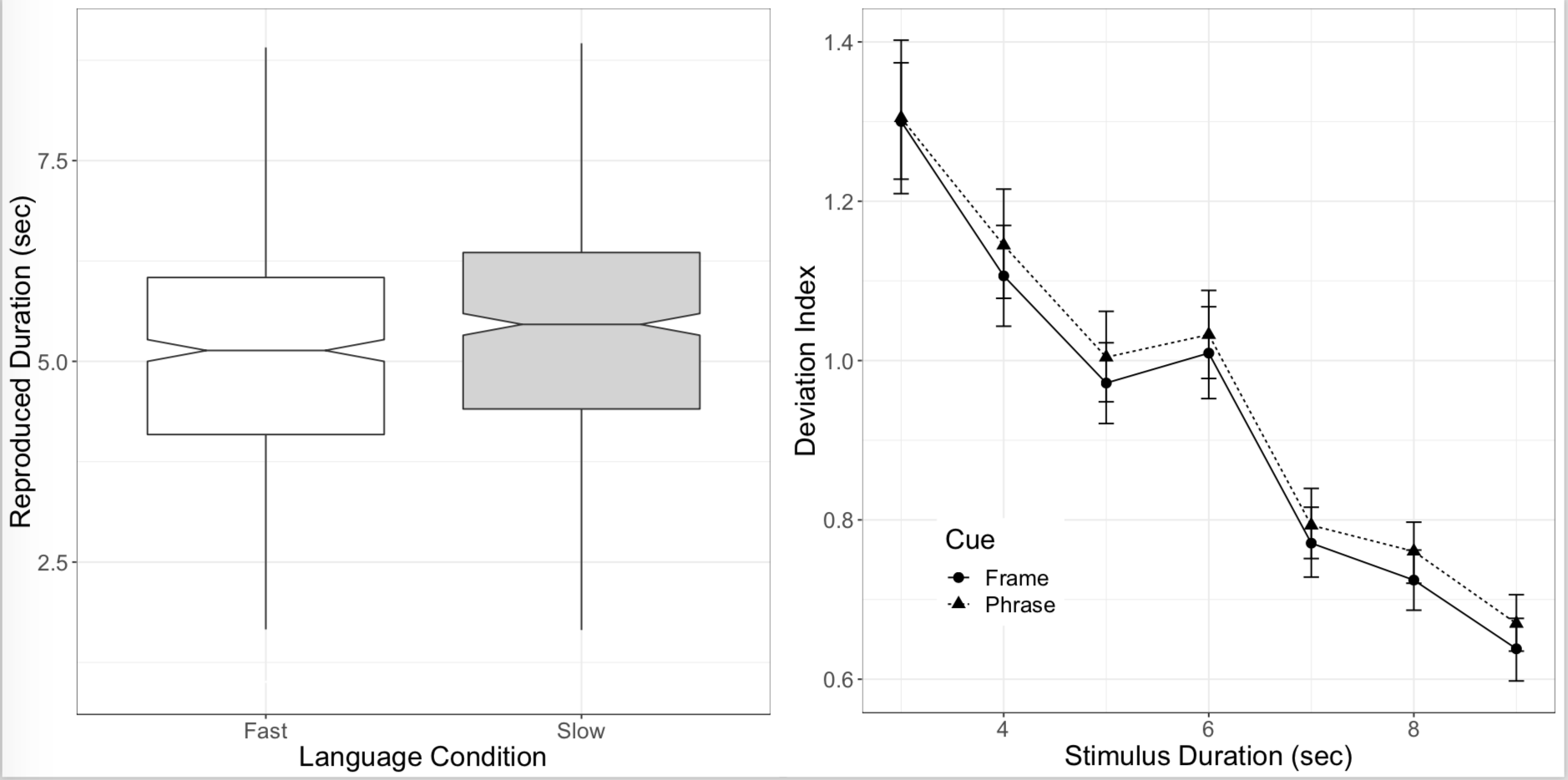


Figure 6: Results of Experiment 3. Panel A shows the reproduced duration as a function of language condition. Notches indicate confidence intervals. Panel B shows the deviation index collapsed across language conditions as a function of stimulus duration and cue type, i.e., Experiments 1 and 3.

**Discussion**

The results of the present experiment are consistent with our previous findings in showing a relationship between the number of words used in the recall task and the reproduced durations, even though the cue phrase constrained the recalled information that could be provided. Importantly, the information retrieved about the event was combined with the conceptual information carried by the phrases, thereby leading to biased longer or shorter reproductions as a function of language. This result is consistent with an interactive retrieval account according to which the memory representation retrieved is combined with top-down conceptual information present at retrieval, leading to a biased reproduction. To our knowledge, this is the first clear demonstration of hybrid event memory representations resulting from language retrieval biases.

One possible explanation for these findings, as discussed in the introduction, is that participants partly relied on the information conveyed by the language because they encoded gist-like characteristics and were uncertain about specific details. Event reproductions were then partially aligned with the phrase concepts to *fill in* the details that were not encoded or retrieved. The effect of language that we have observed may have thus been confounded by relatively weak or sparse speed-related episodic details. Therefore, if the role of language in shaping retrieval is independent of memory strength or its richness, a language effect should still be observed when event information has been more deeply encoded and more event information is learned. We test this possibility next.

**Experiment 4: event reproductions cued by language after several stimulus viewings**

Experiment 4 tested whether the role of language in reproduced duration is independent of encoding conditions. In experiment 3, uncertainty about episodic details could have made participants particularly prone to retrieval biases. Will participants be influenced by language when animations have been learned more deeply? Experiment 2 indicated that event reproductions increased in accuracy with more learning and it is possible that better-learned memories are less susceptible to retrieval biases. Alternatively, language may still exert a modulatory influence, thus suggesting that uncertainty about episodic details does not necessarily drive language effects. Regardless of language influences, we also expect to replicate previous findings concerning the role of recalled information. In particular, if the recalled information is combined with language during event reproductions, the number of words used in event recall should be a significant predictor of reproduced duration in addition to language. Similarly, an increase in reproduction accuracy should be observed when comparing Experiment 3 and 4 due to deeper learning, particularly for longer events in the stimulus set.

**Methods**

**Participants**

58 English native speakers from the University of York who did not participate in previous experiments were recruited for this study. Eight participants were excluded either because their data had more than 30% of extreme values or because they did not provide recall information. There were thus 50 participants in total (25 in each list).

**Design, procedure and statistical analyses**

Stimulus, design and procedure were as in Experiment 2. Participants watched the animations and phrases three times in three different cycles in random order. As in Experiment 3, the phrases were used to cue mental reproductions and verbal recall. Data treatment and analyses were as in previous studies. Due to lack of convergence in the comparison across experiments, the full maximal random effects structure was not possible for deviation indices. See data files. As in Experiment 3, recall accuracy was very high, with 99% of responses providing correct details about the animations (range: 90-100%).

**Results**

Statistical comparisons assessing the effect of language indicated a significant main effect for the reproduced duration (χ2(1)= 10.03, p=0.001) and the deviation index (χ2(1)= 9.16, p=0.002). The effect of language was similar across stimulus duration as there was no interaction between language and stimulus duration. On average, reproduced durations were 308ms longer in the Slow condition than in the Fast condition (mean Slow condition = 5934ms; mean Fast condition = 5626ms), and the deviation index similarly increased or decreased as a function of language (see Figure 7A). Further tests assessing the role of recalled information also indicated significant contributions of word counts and the number of words per second in explaining reproduced durations (χ2(1)= 5.73, p=0.02) and deviation indices (χ2(1)= 9.96, p=0.002), over and above the contribution of stimulus duration and number of segments (see Table 8). Further exploratory comparisons indicated that adding interactions to these models did not increase their fit. These results replicate previous findings and indicate that even though there was more learning, the linguistic concepts modulated event reproductions together with the episodic information recalled.

Table 8: Model summaries for Results of Experiment 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Fixed effects | Estimated coefficient | Standard error | t-value |
| Reproduced | (Intercept) | 1985.02 | 562.91 | 3.53 |
| Duration (ms) | Stimulus duration | 412.33 | 58.87 | **7.00** |
|  | Language condition | -318.41 | 98.09 | **-3.25** |
|  | Stimulus segments | 217.99 | 70.93 | **3.07** |
|  | Log(words) | 641.17 | 256.42 | **2.50** |
| Deviation | (Intercept) | 0.63 | 0.08 | 8.32 |
| Index | Language condition | -0.06 | 0.02 | **-3.11** |
|  | Stimulus segments per second | 0.31 | 0.08 | **3.99** |
|  | Log(words) per second | 0.83 | 0.20 | **4.18** |

Note: Bold values indicate significant fixed effects (p < .05).

**Comparisons across experiments 3 and 4: exposure effect**

As reported for Experiment 1 and 2, the amount of information recalled increased with more learning. There was a main effect of Exposure in explaining log-word count (χ2(1)= 9.85, p=0.002) and no interaction with stimulus duration (see Table 9). Moreover, models assessing whether there were longer or more accurate reproductions for longer animations as a function of exposure revealed a significant interaction between stimulus duration and exposure for reproduced duration (χ2(2)= 8.14, p=0.004) but only a main effect of exposure for the deviation index (χ2(2)= 7.85, p=0.005), see Table 10 and Figures 7B and 7C. The interaction between Exposure and stimulus duration replicates previous results but this interaction did not reach significance for deviation indices, showing only a main effect, despite longer animations being reproduced as longer and as more accurately than in Experiment 3, as shown in Figure 7C. Visual inspection of Figure 7 suggests that the combination of more learned information and language biases might introduce additional distortions compared to Experiment 3, although this possibility requires further investigation. Nevertheless, the present results are generally consistent with those reported above in suggesting a role for recalled information as a function of learning, even though this information was additionally biased by the conceptual information conveyed by language.

Table 9: Mean number of words produced in the recall task for Experiments 3 and 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Stimulus | Experiment 3 | | | Experiment 4 | |
| Duration (sec) | Mean (SD) | | | Mean (SD) | |
| 3 | 15.77 | (8.39) | 21.55 | | (11.08) |
| 4 | 22.81 | (12.41) | 27.10 | | (15.12) |
| 5 | 23.09 | (11.98) | 31.07 | | (17.42) |
| 6 | 27.25 | (17.31) | 37.77 | | (22.86) |
| 7 | 22.14 | (12.57) | 28.77 | | (14.62) |
| 8 | 30.76 | (15.14) | 36.97 | | (20.11) |
| 9 | 26.44 | (15.66) | 31.67 | | (18.82) |

Note: Standard deviations are given in parentheses

Table 10: Model summaries for comparisons across Experiments 3 and 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Fixed effects | Estimated coefficient | Standard error | t-value |
| Reproduced | (Intercept) | 3378 | 615 | 5.49 |
| Duration (ms) | Stimulus duration | **225** | **90** | **2.50** |
|  | Language condition | **310** | **59** | **5.25** |
|  | Exposure | -311 | 311 | -1.00 |
|  | Exposure × stimulus duration | **128** | **44** | **2.93** |
| Deviation Index | (Intercept) | 1.53 | 0.17 | 8.59 |
|  | Stimulus duration | **-0.09** | **0.01** | **-8.93** |
|  | Language condition | **0.05** | **0.01** | **5.15** |
|  | Exposure | **0.01** | **0.10** | **2.86** |

Note: Bold values indicate significant fixed effects (p < .05).

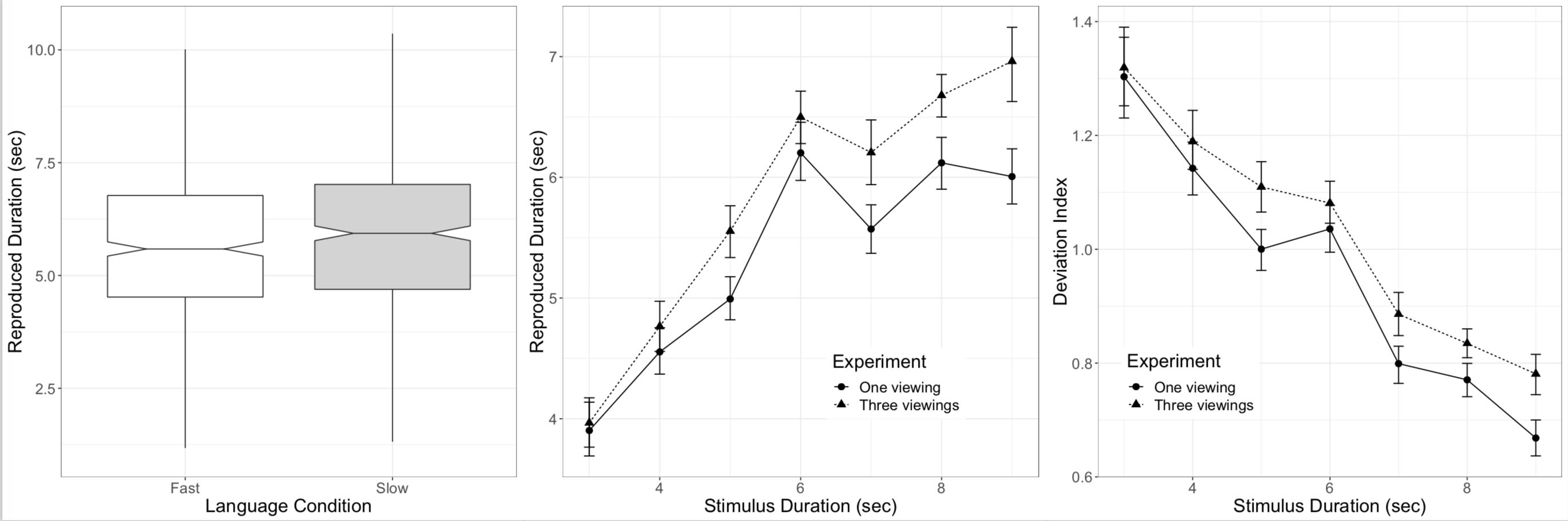


Figure 7: Results of Experiment 4. Panel A shows the reproduced duration as a function of the language conditions. Notches show confidence intervals. Individual scores were adjusted according to Loftus and Mason (1994). Panels B and C show the reproduced duration and the deviation index as a function of stimulus duration and exposure (Experiments 3 and 4). Error bars indicate standard errors computed according to Loftus & Masson (1984) for each experiment separately.

**Discussion**

The results of Experiment 4 indicated a modulatory effect of language, even though participants had the opportunity to study the animations in detail. This result suggests that the role of language is not restricted to situations in which people are uncertain about speed or motion features. Participants could indeed recall many details about the animations, as indicated by the number of words used in the recall task compared to Experiment 3. Despite better learning, the linguistic concepts led to instantiate the unfolding of the events in ways that were consistent with the typical manner of motion entailed by those concepts. Nevertheless, the amount of information recalled also played a role in event reproductions, as evidenced by the predictive role of word counts and the effect of exposure, which were similar in nature to those found in Experiments 1 and 2. Taken together, these results suggest that the episodic details retrieved from memory were combined with the linguistic concepts, to gave rise to event reproductions consistent with both episodic memory and language.

**General Discussion**

The present studies aimed to investigate the relationship between language, event memory and clock time. The studies specifically examined how event reproductions from memory were modulated by event descriptions and the event information recalled. In all studies, descriptive phrases varying in their implied motion speed were associated with the same stimulus animations during learning. Experiments 1 and 2 used a video frame to elicit event reproductions and event recall, and varied the extent to which animations were learned (one viewing vs three viewings). The results of these experiments indicated that event reproductions did not vary as a function of the associated description, suggesting that language did not modulate the way the animations were encoded or subsequently retrieved. Instead, event memory was the main source of information guiding event reproductions, as evidenced by the predictive role of the number of words used in event recall, over and above stimulus duration. Critically, when more episodic details were encoded due to increased stimulus exposure, event reproductions became more accurate, particularly for the longer and shorter animations in the stimulus set. These results are consistent with the recall-based view of memory for duration in that the amount of information learned and subsequently recalled modulated the accuracy and length of event reproductions.

Experiments 3 and 4 investigated whether the use of linguistic cues to retrieve previously seen animations modulated event reproductions, and whether this modulation varied with learning. The results of these experiments indicated that language-mediated retrieval led to shorter or longer reproductions consistent with the phrases, even after extensive learning. Nevertheless, the number of recall words significantly predicted event reproductions. As in Experiment 2, increased stimulus exposure led to more accurate and longer reproductions for the longer animations in the set. The concurrent influence of recalled information and language, therefore, suggests that the retrieved episodic event representations were combined with the information carried by the phrases, leading to hybrid event reproductions modulated by both event memory and linguistic information.

Across all experiments, shorter stimuli tended to be lengthened and longer stimuli tended to be shortened, although this pattern was further modulated by learning and language cues. As expected by the recall-based view, the deviation index in all studies was explained by the density of the details recalled relative to the stimulus duration (the number of words recalled per second), in addition to segmental density (the number of segments per second). This result suggests that the tendency to lengthen short stimuli and shorten long ones may stem from the amount of information recalled, which is proportionally larger for short stimulus events. Taken together, these results are consistent with both a recall-based view of memory for duration and an interactive retrieval account of the role of language in memory. Below we discuss the implication of these findings for current theories.

**Memory and language**

The concurrent influence of recalled information and language in Experiments 3 and 4 suggests that during recollection, episodic event representations were combined with conceptual cues available in the retrieval context, consistent with the interactive retrieval account. This finding contrasts with those of previous studies using verbal labels at retrieval (Hanawalt & Demarest, 1939; Burt & Popple, 1996; Loftus & Palmer, 1974) in that language was present at encoding *and* retrieval so that language did not provide new event information at retrieval. Note that due to the unfamiliar nature of the moving objects in our visual stimuli, these objects had to be linked to their descriptions at encoding. If this association was not already learned, participants would not have been able to retrieve the corresponding geometric figures being referred to by the descriptions at retrieval. For example, when prompted by *athletes in a hurdle race*, they would not know which animation this description referred to (cf. Figure 1). For this reason, we can infer that in Experiments 3 and 4, verbal cues were indeed linked with the animations during learning and these cues mediated access to the corresponding episodic event representation.

Language-mediated retrieval resulted in biased event reproductions likely because the descriptions shaped the nature of the episodic details retrieved via top-down modulations. Phrase cues, unlike cue-frames, constrained participants to retrieve all relevant visual information. The visual aspects of the objects and scenes (e.g., the color and number of geometric figures) and the specific way in which they moved across the scene had to be reconstructed from scratch. This reconstruction process was guided by the phrasal concepts, which provided the overarching schema or gist linking the episodic information being retrieved. Some memory accounts have indeed argued that schema-guided recollection results in episodic and conceptual features being merged, perhaps due to difficulty in discriminating between them at the point of retrieval, even if they were not merged at encoding (Alba & Hasher, 1983; Pezdek et al., 1988; Neuschatz et al., 2002). In this respect, our results provide the first clear demonstration that episodic event features and conceptual features were indeed combined. This finding highlights the possibility that language may modulate retrieval of verbally encoded events in other types of duration judgments and in event memory more generally.

The question now arises as to why language did not modulate encoding or retrieval in Experiments 1 and 2, if access to the true time course of events was not entirely accurate, particularly after a single stimulus exposure. Surely, the linguistic phrases provided familiar information according to which segmental predictions were made and links to existing knowledge were established during learning, thus enabling interactive encoding or retrieval of motion features. One possibility is that episodic and linguistic information are generally encoded as distinct pieces of information and remain so as long as episodic representations can be accessed by means other than language. Language influences on event memory are thus not observed unless retrieval is mediated by language. This possibility is consistent with additional studies that we have conducted showing that language effects with the present stimuli emerge after a period of memory consolidation irrespective of retrieval cue, for example, a day after encoding (Wang, Gaskell, & Gennari, 2018). This finding indicates that memory consolidation might be necessary for linguistic and visual information to be merged in memory, a result consistent with current models of memory consolidation (McClelland, 2013; Moscovitch, Cabeza, Winocur, & Nadel, 2016; van Kesteren, Ruiter, Fernández, & Henson, 2012; Walker & Stickgold, 2010).

Taken together, our results have implications for theories on the relationship between language and memory. Prior research has suggested that linguistic categorization of scenes, objects or colors may lead to distorted encoded representations that impair subsequent recognition of within-category instances (Feist & Gentner, 2007; Lupyan, 2008, 2012; Regier & Kay, 2009; Roberson, Davidoff, Davies, & Shapiro, 2005). Although these studies have examined different types of stimuli that may involve disparate processes, our results suggest that language effects may stem from retrieval mechanisms, rather than distorted encoded representations. Under this view, language impairs within-category discrimination because labels are implicitly or explicitly used to access (often difficult to retrieve) episodic features, and thus, retrieval is mediated by language. Language-induced retrieval may particularly operate in situations in which language has been explicitly or implicitly used at encoding to discriminate among highly similar stimuli, thus priming or re-activating the labels upon seeing previously labeled stimuli. This might explain why some color label effects disappear when labeling is prevented by task demands (Winawer et al., 2007). Unlike objects and colors, however, dynamic events contain multiple features and dimensions (space, motion, actors, goals) that are less likely to necessitate discriminating labels at encoding. On the contrary, the presence of multiple episodic event features is likely to increase distinctiveness between stimulus events and thus may lead to episodic and conceptual features being separately encoded rather than merged, as observed here. The question therefore remains as to whether language-induced conceptualizations are generally able to augment or distort encoded representations rather than retrieved ones, particularly in the domain of dynamic events.

**Memory for time and its biases**

Previous research on memory for duration has mostly focused on unfamiliar stimuli and has manipulated stimulus properties (Avni-Babad & Ritov, 2003; Block & Reed, 1978; Boltz, 1995, 1998, 2005, Faber & Gennari, 2015b, 2015a; Ornstein, 1969; Poynter, 1983, 1989; Zakay et al., 1994). This research has shown that the number of segments, their hierarchical structure and their predictability modulate duration judgments. In contrast, we have investigated familiar events and focused on recollection processes. Although our measure of segmentation was obtained in a separate off-line task and therefore did not represent an individual’s on-line segmentation, segmentation was also a good predictor of reproduced duration in all experiments. This was expected from the fact that event segmentation and event memory are closely related, with more segments leading to more information recalled (Zacks et al., 2007; Hanson & Hirst, 1994). The present results, therefore, extend Ornstein’s and event structure approaches to duration memory by demonstrating that the information actually recalled about an event plays a significant role in explaining reproductions of familiar events.

In all our studies, there was a tendency to lengthen short stimulus events and to shorten long ones, despite additional influences of learning and language. This bias in duration reproductions was partially accounted for by independent measures of recalled information, in particular, the density of the information recalled relative to the stimulus duration. This sort of bias—like most biases observed in connection with segmental structure— can be obtained when participants judge a single event sequence, as in naturalistic situations such as witness testimony (Roy & Christenfeld, 2008; Yarmey, 2000; Tobin et al., 2010; Ornstein, 1069; Jeunehomme et al, 2017; Block, 1978, 1982; Zakay, 1993; Zakay et al., 1994). Although learning influences that we have not measured may also operate in our studies, these previous results and the present findings indicate that segmental structure, and more generally the density of the information recalled, play a substantial role in explaining memory-based estimation biases. A key question to understand these biases therefore is why more information is proportionally recollected for short stimuli than long ones.

A growing body of evidence suggests that the density of information recalled may stem from event segmentation and its consequences for recollection. It has been shown that perceptual features at event boundaries are recognized and recalled better than those occurring within boundaries (Boltz, 1992; Hanson & Hirst, 1989; Newtson & Engquist, 1976; Schwan & Garsoffky, 2004; Swallow et al., 2009), and that similar segments can interfere with each other at retrieval, resulting in impaired memory (Horner, Bisby, Wang, Bogus, & Burgess, 2016; Radvansky & Copeland, 2006; Radvansky, Tamplin, & Krawietz, 2010; Radvansky, Zwaan, Federico, & Franklin, 1998; Radvansky & Zacks, 2017). In the real world, as in our stimulus animations, shorter events tend to have fewer segments and fewer episodic details than longer ones, as shown by positive correlations between the words recalled and stimulus duration (Tables 2 and 5). In longer events, therefore, there are more opportunities to miss or forget within-event perceptual information and/or to experience interference at retrieval. Within-event information is critical for accurately mapping the event representation onto real time, particularly for long segments. For example, the accurate duration reconstruction of a slow and smooth motion across the screen not only depends on the event boundaries recalled (e.g., path changes or places encountered during travel), but also on what could be reconstructed to have occurred in between these boundaries. Thus, in longer events, there might be more forgetting of within-segment information or more difficult retrieval when segmental features are relatively similar. For short events, in contrast, fewer and shorter segments and their details are more fully encoded and retrieved, creating a relatively crowded event representation in which many things happened, hence the lengthened reconstructions.

Another possible mechanism operating at segmentation is coarse hierarchical structuring based on prior knowledge, which might be particularly relevant for long events: the length of the visual stimulation itself may lead to higher levels of structuring as more information accrues with increased clock duration, resulting in shorter reproductions for stimuli that are more schematically encoded. Consistent with this view, stimuli that are organized into schematic representations due to being repetitive or familiar are judged as shorter than variable or unfamiliar stimuli of the same duration, indicating that schematization leads to temporal compression (Ornstein, 1969; Boltz, 1998; Avni-Babad & Ritov, 2003; Faber & Gennari, 2015). Similarly, long movies or narratives are organized into higher-order structures according to agents’ goals and causal links (Bauer, 1992; Lichtenstein & Brewer, 1980; Loucks, Mutschler, & Meltzoff, 2017; Shank & Abelson, 1977; Zacks, Tversky, & Iyer, 2001; Zwaan & Radvansky, 1998). This organization often enables the recollection of the main story line in a compressed or schematic form but not necessarily of every detail. These observations therefore suggest that as events become longer, they tend to be compressed into more schematic representations that work against accuracy in duration reproductions.

These mechanisms are consistent with the specific pattern of reproduction biases that we have observed. High-density representations for short events were reconstructed as highly eventful, leading to longer reproductions. On the other hand, the animations with the largest number of segments and the highest segmental or word density were five- and six-second animations (see Tables 2 and 5). This may have boosted reproduction accuracy for these animations relative to longer animations with fewer segments (cf. Figure 4), leading to better alignment with clock time during mental reproductions, i.e., retrieved representations were not too crowded or too sparse relative to clock duration. Finally, longer animations in the set tended to have fewer but longer segments that were also similar in nature, e.g., shapes swimming in a tank, a square traveling on a path, or a circle going down a staircase. Aspects of within-segment information were likely schematized, forgotten or difficult to retrieve, thus leading to the sharp decrease in accuracy observed for seven-second animations relative to five- or six-second animations. In fact, as shown in Figure 2, the reproduced duration of seven-to-nine second animations did not differ by much after one viewing, consistent with the possibility that stimulus information loss increases as segments become longer.

In sum, we suggest that the density of the information recalled relative to stimulus duration, which we have shown modulates reproduced duration, derives from event segmentation and retrieval mechanisms, consistent with event structure approaches to memory for time. Although further research on the specific memory mechanisms at hand is needed, information density provides new insights to understand long-standing issues on temporal memory biases and the representation of time in memory. The way our minds organize fleeting experiences into events largely determines how the duration of past experiences will be reconstructed, thus leading to time representations that do not necessarily coincide with clock time. Nevertheless, these event mechanisms are adaptive in that they allow the human brain to process and efficiently store relevant information from dynamic experiences.

Acknowledgments: We thank Dr Marcus Harrington for kindly reading and commenting on the manuscript.

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. https://doi.org/10.1002/hipo.20405

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231.

Avni-Babad, D., & Ritov, I. (2003). Routine and the perception of time. *Journal of Experimental Psychology. General*, *132*(4), 543–550. https://doi.org/10.1037/0096-3445.132.4.543

Bailey, N., & Areni, C. S. (2006). When a few minutes sound like a lifetime: Does atmospheric music expand or contract perceived time? *Journal of Retailing*, *82*(3), 189–202. https://doi.org/10.1016/j.jretai.2006.05.003

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Bauer, P. J. (1992). Holding it all together: How enabling relations facilitate young children’s event recall. *Cognitive Development*, *7*(1), 1–28. https://doi.org/10.1016/0885-2014(92)90002-9

Block, R. A. (1978). Remembered duration: Effects of event and sequence complexity. *Memory & Cognition*, *6*(3), 320–326. https://doi.org/10.3758/BF03197462

Block, R. A. (1982). Temporal judgments and contextual change. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *8*(6), 530–544. https://doi.org/10.1037/0278-7393.8.6.530

Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 656.

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, *4*(2), 184–197. https://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.

Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics*, *57*(7), 1080–1096. https://doi.org/10.3758/BF03205466

Boltz, M. G. (1998). Task predictability and remembered duration. *Perception & Psychophysics*, *60*(5), 768–784.

Boltz, M. G. (2005). Duration judgments of naturalistic events in the auditory and visual modalities. *Perception & Psychophysics*, *67*(8), 1362–1375. https://doi.org/10.3758/BF03193641

Burt, C. D., & Popple, J. S. (1996). Effects of implied action speed on estimation of event duration. *Applied Cognitive Psychology*, *10*, 53–63.

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived forms. *Journal of Experimental Psychology*, *15*, 73–86.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology*, *20*(3), 396–413. https://doi.org/10.1080/09541440701728623

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159.

Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in language comprehension. *Cognitive Psychology*, *62*(1). https://doi.org/10.1016/j.cogpsych.2010.09.002

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2012). Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Attention, Perception, and Psychophysics*, *74*(8), 1819–1841. https://doi.org/10.3758/s13414-012-0362-4

Faber, M., & Gennari, S. P. (2015a). In search of lost time: Reconstructing the unfolding of events from memory. *Cognition*, *143*, 193–202. https://doi.org/10.1016/j.cognition.2015.06.014

Faber, M., & Gennari, S. P. (2015b). Representing time in language and memory: The role of similarity structure. *Acta Psychologica*, *156*, 156–161. https://doi.org/10.1016/j.actpsy.2014.10.001

Faber, M., & Gennari, S. P. (2017). Effects of Learned Episodic Event Structure on Prospective Duration Judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *published*. https://doi.org/10.1037/xlm0000378

Feist, M. I., & Cifuentes Férez, P. (2013). Remembering How: Language, Memory, and the Salience of Manner. *Journal of Cognitive Science*, *14*, 379–398.

Feist, M. I., & Gentner, D. (2007). Spatial Language influences memory for spatial scenes. *Memory and Cognition*, *35*(2), 283–296.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role. *Journal of Memory & Language*, *44*(4), 516–547.

Gennari, S. P. (2004). Temporal references and temporal relations in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*(4), 877–890. https://doi.org/http://dx.doi.org/10.1037/0278-7393.30.4.877

Gennari, S. P., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, *89*(1), B27–B41. Retrieved from http://dx.doi.org/10.1016/S0010-0277(03)00069-6

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*(1), 49–79.

Gentner, D., & Loftus, E. F. (1979). Integration of Verbal and Visual Information as Evidenced by Distortions in Picture Memory. *The American Journal of Psychology*, *92*(2), 363–375.

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.

Hanawalt, N. G. (1937). Memory trace for figures in recall and recognition. *Archives of Psychology*, *216*, 1–81.

Hanawalt, N. G., & Demarest, I. H. (1939). The effect of verbal suggestion in the recall period upon the reproduction of visually perceived form. *Journal of Experimental Psychology*, *25*, 159–174.

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2525593

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, *154*, 151–164. https://doi.org/10.1016/j.cognition.2016.05.013

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590

Jeunehomme, O., Folville, A., Stawarczyk, D., Van der Linden, M., & D’Argembeau, A. (2017). Temporal compression in episodic memory for real-life events. *Memory*. https://doi.org/https://doi.org/10.1080/09658211.2017.1406120

Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology*, *12*(3), 412–445. https://doi.org/10.1016/0010-0285(80)90015-8

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobil destruction: an example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior*, *13*, 585–589.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. https://doi.org/10.3758/BF03210951

Loucks, J., Mutschler, C., & Meltzoff, A. N. (2017). Children’s Representation and Imitation of Events: How Goal Organization Influences 3-Year-Old Children’s Memory for Action Sequences. *Cognitive Science*, *41*(7), 1904–1933. https://doi.org/10.1111/cogs.12446

Lupyan, G. (2008). From Chair to “Chair”: A Representational Shift Account of Object Labeling Effects on Memory. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.137.2.348

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*(MAR), 1–13. https://doi.org/10.3389/fpsyg.2012.00054

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. https://doi.org/10.1037/a0033812

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annual Review of Psychology*, *67*(1), 105–134. https://doi.org/10.1146/annurev-psych-113011-143733

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436–450. https://doi.org/10.1016/0022-1031(76)90076-7

Ornstein, R. E. (1969). *On the experience of time*. Harmondsworth, England: Penguin.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155–184. https://doi.org/10.1016/j.cognition.2008.02.007

Poynter, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition*, *11*(1), 77–82.

Poynter, W. D. (1989). Chapter 8 Judging the Duration of Time Intervals: A Process of Remembering Segments of Experience. *Time and Human Cognition: A Life-Span Perspective*, *59*(C), 305–331.

Prentice, W. C. H. (1954). Visual Recognition of Verbally Labeled Figures. *The American Journal of Psychology*, *67*(2), 315–320.

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: situation models and experienced space. *Memory & Cognition*, *34*(5), 1150–1156. https://doi.org/10.3758/BF03193261

Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychological Bulletin Review*, *17*(6), 900–904. https://doi.org/10.3758/PBR.17.6.900

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 608–620. https://doi.org/10.1002/wcs.133

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioural Sciences*, *17*, 133–140. https://doi.org/10.1016/j.cobeha.2017.08.006

Radvansky, G. A., Zwaan, R. A., Federico, T., & Franklin, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 1224–1237.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446. https://doi.org/10.1016/j.tics.2009.07.001

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001

Rock, I., & Engelstein, P. (1959). A Study of Memory for Visual Form Author. *The American Journal of Psychology*, *72*(2), 221–229.

Roy, M. M., & Christenfeld, N. J. S. (2008). Effect of task length on remembered and predicted duration. *Psychonomic Bulletin and Review*, *15*(1), 202–207. https://doi.org/10.3758/PBR.15.1.202

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., … Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71. https://doi.org/10.1016/0010-0285(90)90003-M

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology*, *18*(1), 37–55. https://doi.org/10.1002/acp.940

Shank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.

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. https://doi.org/10.1037/a0015631

Tobin, S., Bisson, N., & Grondin, S. (2010). An ecological approach to prospective and retrospective timing of long durations: A study involving gamers. *PLoS ONE*, *5*(2), 16–18. https://doi.org/10.1371/journal.pone.0009271

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2010.02.006

Tulving, E. (1984). Précis of Elements of episodic memory. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X0004440X

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Cognitive Sciences*, *35*(4), 211–219.

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: sleep-dependent memory evolution. *Nature Reviews. Neuroscience*, *11*(3), 218; author reply 218. https://doi.org/10.1038/nrn2762-c1

Wang, Y., Gaskell, G., & Gennari, S. . (2018). Sleep-dependent integration of episodic events into language-induced schemas. *In Prep.*

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

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

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. https://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.

Zakay, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics*, *54*(5), 656–664. https://doi.org/10.3758/BF03211789

Zakay, D., Tsal, Y., Moses, M., & Shahar, I. (1994). The role of segmentation in prospective and retrospective time estimation processes. *Memory & Cognition*, *22*(3), 344–351.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–182. https://doi.org/10.1037/0033-2909.123.2.162

|  |  |
| --- | --- |
| **Supplemental Materials**  **1. Stimulus phrases and cue frames** | |
| *A firework rocket launched into the sky*  *A Chinese lantern raising up into the sky* | *Planes flying in the sky*  *People paragliding in the sky* |
|  |  |
| (3 sec., 2.2 segments) | (3 sec., 3.3 segments) |
|  |  |
| *A racing car going downhill.*  *An army tank going downhill* | *Cutting a banana with a knife*  *Cutting a log with a handsaw* |
|  |  |
| (3 sec., 2.2 segments) | (4 sec., 4.1 segments) |
|  |  |
| *Someone gulping a glass of beer.*  *Someone drinking a glass of beer* | *Products in a factory's assembly line*  *People lining up to buy food in a canteen* |
|  |  |
| (4 sec., 3.2 segments) | (4 sec., 3.5 segments) |
|  |  |
| *A monkey climbing a tree to pick a fruit*  *An old man climbing a tree to pick a fruit* | *Children racing in a school yard*  *Horses sprinting ar a race course* |
|  |  |
| (5 sec., 3.6 segments) | (5 sec., 7.1 segments) |
|  |  |
|  |  |
| *A speed boat crossing a river*  *A man swimming across a river* | *Two friends jogging to meet in the park*  *Two friends strolling to meet in the park* |
|  |  |
| (5 sec., 5.9 segments) | (6 sec., 4.4 segments) |
|  |  |
| *Cyclists in a relay competition*  *Students in a relay competition* | *Athletes competing in a hurdle race*  *Army cadets crawling under wire obstacles* |
|  |  |
| (6 sec, 4 segments) | (6 sec, 5.7 segments) |
|  |  |
| *Kayakers competing in the river*  *Visitors rowing in the river* | *A child running downstairs to find mom*  *A child crawling downstairs to meet mom* |
|  |  |
| (7 sec., 2.6 segments) | (7 sec., 6.6 segments) |
|  |  |
| *A person skating on a frozen lake*  *A person stepping on to a frozen lake*. | *Mixing ingredients in the bowl*  *Simmering vegetables in the pot* |
|  |  |
| (7 sec., 2.9 segments) | (8 sec., 5 segments) |
|  |  |
| *Taking the tube home*  *Taking the school bus home* | *Students running up to their classroom*  *Students walking up to their classroom* |
|  |  |
| (8 sec., 4.1 segments) | (8 secs, 5.7 segments) |
|  |  |
| *An ambulance taking someone to the hospital.*  *Grandma taking the bus to the hospital* | *A tourist driving a car on a mountain road*  *A tourist riding a mule on a mountain path* |
|  |  |
| (9 sec., 6.2 segments) | (9 sec., 2.4 segments) |
|  |  |
| *Dolphins swimming in an aquarium pool*  *Goldfish swimming in an aquarium tank* |  |
|  |  |
| (9 sec., 3.8 segments) |  |
|  |  |

**2. Stimulus Pre-tests**

Participants in all pre-test studies were recruited from Prolific Academic, a UK-based research site. All pre-tests studies used a similar design: the 42 phrase-animation pairs were arranged in two lists with 10 or 11 fast and slow phrases in each list. The two phrases associated with an animation were assigned to different lists so that each animation was seen only once in a list.

**Implied motion speed***.* To examine whether the two phrases associated with an animation indeed implied events of different speed, we asked native English participants (N = 48) to judge the typical pace of the events described by phrases on a scale of 1 to 7 (1= very slow pace, 7= very fast pace). Participants were instructed to use their general knowledge about the events described and examples were provided referring to the extreme points of the scale (e.g., *a snail crawling on a leaf* vs *a missile flying through the sky*). Means and standard deviation are shown in Table 1. An independent-samples t-test across phrases revealed that phrases in the fast condition were rated as faster than the phrases in the slow condition (t(40)=-6.82, p<.001). This result thus demonstrates that the phrases used in our studies indeed implied different motion speeds.

**Phrase-animation fit**. Because each animation was to be described with a fast or slow phrase, the actual speed in the animation had to be a possible speed of both the fast and slow moving object being described. Successive questionnaires during stimulus creation informed our choice of animation speed. In a final questionnaire, we checked that the two phrases associated with an animation were equally good descriptors, i.e., the phrases fit the animations equally well. The questionnaire asked native English speakers (N= 32) to rate how well the descriptive phrase fitted the animation on a scale from 1 (does not fit at all) to 7 (fits very well), assuming that the depicted objects represented those named in the phrase. Means and standard deviation are displayed in Table 1. A paired-sample t-test across animations revealed no significant difference between fast and slow phrases (t(20)=-1.60, p>.05). Overall ratings fell within a 4-6.5 range, indicating that the language labels fit the animations well.

**Event familiarity**. In this test, we checked that the familiarity of the events described by the phrases did not differ across conditions. In a familiarity rating questionnaire, participants (N=63) were instructed to rate how familiar they were with the events described on a scale from 1 (not familiar at all) to 7(very familiar). Familiarity was defined as the degree to which they came in contact with the event by seeing it, hearing about it or participating in it and examples were provided (*Friends having dinner* vs. *an elephant stomping on a doll*). Stimulus phrases were assigned to two lists as indicated above. See descriptive statistics in Table 1. An independent-sample t-test across phrases revealed no significant difference in familiarity between phrases in the two speed conditions (t(40)=-1.61, p>.05).

**Perceived scene scale**. When perceiving the animations, the phrases prompted some observers to view the image as closer or further away. For example, the birds’ eye view of the city streets in the last example of Figure 1 can be perceived as further away if the object is moving slowly. To check whether the language condition was systematically related to the scale with which the image was perceived, we devised a questionnaire in which participants (N=36) were required to judge the relative scale of the scene in the animation on a scale of 1 to 7. Animation-phrase pairs were used as stimuli. Example of extreme cases were provided (large scale: *cruises traveling between islands in the ocean* and an aerial image with dot-like islands in an ocean; small scale: close-up image of a snail on a leaf and corresponding phrase). Descriptive statistics are shown in Table 1. A paired-sample t-tests across animations revealed no significant difference between the fast and slow versions (t(20)=-.988, p>.05).

**Event segmentation**. To obtain an average measure of the number of segments perceived in each animation, we conducted a web-based questionnaire (N=62). We asked participants to first preview the animation and the description to understand what happens in the animation, and then count *the number of times that you think a unit of something happening ends and another begins, that is, how many units you can identify in the animations.* Participants were then asked to pay attention to the smallest possible units they could identify and count units with their fingers or marks on a paper. An example was provided showing an animation of a circle moving about in a maze of pipes, which represented a plumber fixing a leak. Possible example units were suggested, which included turning a corner in the maze and moving an object into a pipe. Nevertheless, it was emphasized that there were no correct and incorrect responses, so that participants should use their own judgment in determining when a unit occurred. Participants could replay the animation by clicking on the play button and entered their number on a textbox underneath the animation. The web page recorded the time it took participants to complete the task. From this measure, we can infer that participants on average watched the animations twice (the questionnaire duration roughly doubled the animations’ total duration). The average number of segments across participants was used as a predictor in our analyses.

We also tested our stimuli with previously used segmentation tasks for comparison. In this task (N=58), participants first previewed all animations and descriptions and then performed the segmentation task by clicking the mouse as the animations played. The instructions were similar to those described above. Participants in this task reported finding the task difficult and missing segments. Short animations often did not have any segments. The number of segments reported was also consistently smaller than that of the timed task (the mean number of segments was 4.22 vs. 2.10 respectively), likely due to participants only indicating major segments (coarse segmentation). The two distributions however were highly correlated (Spearman rho=.84). Because our main experiments required careful learning of the animations for later memory tests, we reasoned that untimed segmentation was more likely to represent this learning.

**3. Word counts as a recall measure**

In Experiment 1A, we manually coded participants’ verbal recall to compare this measure to word counts. In this coding, we counted the number of color and movement changes described, as well as the appearance or disappearance of objects on the screen. Each change was coded as 1. Examples of this coding are provided in Table SM1 for the three animation of Figure 1 of the main text. A Spearman correlation across the distribution of word counts and number of changes indicated a strong relationship (rho=.84, p<.0001). This coding however only considered three types of changes in the animation. One may argue that recalling other aspects of the animation is also relevant, for example, the color of objects absent in the probe or action and location qualifiers (*erratically, in their lanes, into the sky*). With such decisions, the number of details would become closer to the number of words. Since function words such as *the* or *and* are unlikely to differ systematically across items, we reasoned that manual coding would add very little to the information provided by the word counts.

|  |  |  |
| --- | --- | --- |
| Animation recall | Word Count | Changes coded |
| *Rectangle turns red and get a little flame underneath. It moves upwards towards the top of the box.* | 18 | 3 |
| *The firework took off and travelled vertically into the sky.* | 10 | 2 |
|  |  |  |
| *The runners of the relay move erratically (but in their lanes) upwards, two more squares enter the frame from the bottom, the left one is blue, the 3rd one along is grey, they overtake the other two runners and the grey one (third runner along) finishes first* | 47 | 7 |
| *The army squares crawl up, two more which are purple and green follow behind them in the first and third columns* | 21 | 4 |
|  |  |  |
| *The uncolored ambulance moves towards the junction, then turns south and left, at the next turn, the red circle moves into the rectangle and travels along past the oval and up towards the middle rectangle on the right, which turns a chrome color and the red circle moves inside it.* | 50 | 9 |
| *The ambulance picks up a patient and takes them to the hospital, which is in the center building on the right* | 21 | 2 |

Table 1: Example of verbal recall in Experiment 1A with coded changes and number of words