Time in Language and Memory

Yaqi Wang

PhD

University of York

Psychology

March 2019

ABSTRACT

This thesis aimed to investigate how we represent event duration in memory and language, and how linguistic concepts modulate event duration representation. In Chapter 2, we investigated how people represent the duration conveyed by temporal adverbials during story comprehension. We found that longer events took longer to process, as more diverse world knowledge was associated with longer events. We argue that duration representation in language comprehension is rooted in activation of world knowledge. In Chapter 3 and Chapter 4, we investigated how people represent event duration from episodic memory and how language shapes event encoding and recollection. We found that the amount of verbally recalled information predicted the remembered event duration. The more recalled information, the longer the remembered duration. This relationship explains that deeper learning lengthens duration reproductions and the Vierordt's law. Moreover, we found that linguistic concepts introduced at encoding do not bias reproduced duration immediately, unless linguistic concepts prompted recollection. However, after 12hs delay with sleep, linguistic concepts biased the remembered duration. This result suggested that episodic details and linguistic concepts were initially stored independently and then integrated during sleep or language-mediated recollection. In Chapter 5, we explored how concepts and event structures modulate event segmentation. We found that concepts lead to coarser segmentation only when they convey goal-specific events rather than goal-unspecific events. We argue that conceptualizing abstract animations into goal-specific events enable us to chunk smaller segments into larger units according to a goal-hierarchy, leading to coarser segmentation. In contrast, goal-unspecific events by nature lack a goal-hierarchical structure. Therefore, verbal concepts contribute more to event segmentation for goal-specific events than goal-unspecific events. Taken together, these results support the information activation-based account for duration representation in language comprehension, and highlight the role of verbal concepts in shaping duration representation and event segmentation.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor Dr. Silvia Gennari for all her support throughout my Ph.D. journey. Her enthusiasm in research inspired my academic interests. Her excellent guidance, caring and patience make me grow to be a real researcher. She is THE MENTOR of my academic study and research!

In addition, I am also hugely appreciative to Professor Gareth Gaskell for his guidance and support. I knew him, when I did my lab placement during my master period. I also collaborate with him on a sleep study during my PHD. Besides, he is also a member of my Thesis Advisory Panel. It is great pleasure to work with him. He not only motivated my interests in sleep study, but also helped me with the experiment design and contributed ideas to writing up the sleep study.

Similarly, profound gratitude goes to Professor Sven Mattys who kindly agreed to be my TAP member. It has been a challenge for me to set up a plan for future study, but his encouragement and insightful comments helped me to clear my thoughts and figure out my direction.

In addition, many thanks to my dear friends Xiuyi Wang and Ling Ge. There would be no one with whom I can share my tears and fears, if they were not here. They believed in me even when I stopped believing in myself. Thanks for always giving me reasons to trust myself.

Finally, I would like to thank my parents for their support throughout my PHD. They not only provided me financial supports, but also cheered me up and stood

by me through good times and bad. I would be unable to do a PHD without their support and love. I am so lucky to have parents like them!

DECLARATION

I declare that this thesis is a presentation of my original research work under the supervision of Dr Silvia Gennari, in the Department of Psychology, at the University of York. For the present thesis, which I am submitting to the University, no degree or diploma or distinction has been conferred on me before, either in this or in any other University.

In addition, among all the six chapters in the thesis, Chapter 3 is a published paper, which has been accepted by Cognitive Psychology (see below). I declare that I carried out this study on my own under the guidance and supervision of Dr Silvia Gennari. In addition, Chapter 4 is a publication as well, which has been sent to Journal of Experimental Psychology: General and is under review now. I conducted the experiments on my own with the collaboration of Professor Gareth Gaskell under the supervision of Dr Silvia Gennari's. Specifically, Silvia supervised all the experiments and Gareth contributed ideas in experiment design, data collection, data analysis and writing. Besides, the rest of chapters (Chapter 1, Chapter 2, Chapter 5 and Chapter 6) are all my own original work carried out for the degree of Doctor of Philosophy in Psychology under the guidance of Dr. Silvia Gennari. The data reported in chapters 2, 3, 4, and 5 have been presented at the following conferences and publications. All sources are acknowledged as References.

Publication:

Wang, Y., & Gennari, S. P. (2019). How language and event recall can shape memory for time. Cognitive psychology, 108, 1-21.

Wang, Y., Gaskell, M. Gareth and Gennari, S. P. (2019). Language influences on event memory are mediated by sleep-dependent consolidation. General

Experimental Psychology: General (under review).

Conference:

Wang, Y., & Gennari, S. P. (2016). Event duration for small- and large-scale events language comprehension. *Poster presented at the Architecture and Mechanisms of Language Processing,* Bilbao, Spain.

Wang, Y., & Gennari, S. P (2017). Memory of event duration. *Poster Presentation at the Experimental Psychology Society Meeting,* Reading, UK.

Wang, Y., & Gennari, S. P (2017). Language effects in memory for event duration. *Poster presented at the Architecture and Mechanisms of Language Processing,* Lancaster, UK.

CHAPTER 1

Processing event duration in language and memory

1.1 Overview

As human beings, we get the feeling of time passing by experiencing discrete events. When we talk about a period of time which we have experienced in the past, we usually talk about the events happening in that interval. In an obvious sense, time and events always interact and are bound together. On the one hand, the representation of time is determined by the experience of events occurring in time. On the other hand, we acquire abundant knowledge of event durations and temporal relations between events from the experience of events which we have stored as part of our world knowledge and this in turn influences the representation of an event. For example, when we talk about someone ‘reading a book for five minutes’, we would not usually expect the person to finish reading the whole book because, based on our own experience, it usually takes several hours to read a book. Moreover, in reading and conversation, we use language to talk about events and our words reflect or imply their temporal development (for example, walking, strolling, running) and duration (for example, ten minutes, five days). The temporal information conveyed by language therefore serves as the cues to activate the associated event knowledge, which enables us to represent the unfolding of events over time.

In this thesis, I aim to investigate the representation of event duration in language and episodic memory. In the present chapter, I review the major theories concerning the representation of time in language comprehension and episodic memory. In Chapter 2, I discuss how people represent event duration in language when event duration is explicitly expressed by temporal adverbials. Following language comprehension research reviewed here, I examine the time taken to retrieved previously read information from short stories. This retrieval time is taken to indicate aspects of the event representations constructed during reading, and thus, we used this measure to test predictions concerning the representation of events. In Chapter 3 and Chapter 4, I will move on to duration representation for experienced past events. To be more specific, I explore how people reconstruct event duration from episodic memory and how this process is shaped by language or schematic knowledge. To this end, I examine previously used measures such as that of reproduced duration (i.e., the mental replay of a past event from memory) to infer how people represent the temporal unfolding of previously observed events. Several studies in this chapter manipulate the presence of linguistic descriptions at encoding or retrieval to assess how language may modulate our representations of event duration. In Chapter 4, I discuss how event duration may be modulating by overnight sleep, a process known to alter encoded representations, as reviewed below. Finally, in Chapter 5, I will expand my study further to explore how the internal structure of event indicated by language affects event representation.

Here, I shall summarize and discuss how previous studies and theories have explained the processing of temporal information about events. To be more specific, in section 1.2, I shall introduce the basic theories of event, especially, event duration representation in language. Therefore, this section provides the theoretical background for Chapter 2. In section 1.3, I shall move on to a more detailed literature review about how humans represent time from their real-world experiences, which are closely related to the studies in Chapter 3 and Chapter 4. Additionally, the literature reviews relating to event structure will be presented in section 1.2.3, which motivates our study of event segmentation in Chapter 5. Moreover, in each section, I will also discuss the contributions and limitations of the existing theories or models in explaining the question of interests: duration representations in language and episodic memory.

1.2 Event representations

We all live in a continuous information stream and we make sense of this dynamic experience by segmenting it into discrete meaningful events (Zacks & Tversky, 2001 ; Zacks, 2004). Events are the basic units of our lives, and an event is defined as “a segment of time at a given location that is conceived by an observer to have a beginning and an end” (Zacks & Tversky, 2001). In other words, an event is a cluster of experiences perceived by people in a shared space/time dimension. So how do we manage to parse the ongoing information into discrete events? How do we memorize and retrieve an event from our memory? How do we extract knowledge of time or event duration from our experiences and understand temporal information conveyed by language? In this section, I shall review the literature relating to these questions.

1.2.1 Memory basis for event representation

1.2.1.1 Memory stages

Memory is critical to us and enables us to learn and recall a large amount of information from previous experience. We all possess a large amount of knowledge about the world. For example, what is the capital of the UK? How many days are there in a month? What did we have for breakfast? When was the last time that we have dinner with friends? We can easily find answers to these questions by making use of our memory. How do we manage to memorize the events and how many tipes of memory are there? In this section, I will firstly introduce memory stages and then focus on the two types of memrories we are interested in. Therefore, we can have a general understanding of the memory basis for event representation.

It has been suggested there are three stages involved in memory processes: encoding, storage, retrieval and consolidation (McLeod, 2007; Walker & Stickgold,2004). The initial stage of memory is information encoding. In this stage, we are exposed to perceptual information and input the sensory signals into cognitive/memory system. Thus, memory is initially formed or acquired at encoding stage (Walker & Stickgold,2004).

Following acquisition or encoding, the next stage of memory is information storage, during which the encoded information is stored in memory for later retrieval. In this stage, memory would be consolidated. Classically “memory consolidation” indicates the process whereby a memory becomes stabilized and resistant to disruption or interference (McGaugh, 2000; Walker & Stickgold,2004). To be specific, memories for an event initially retain highly detailed perceptual information. But over time, only fewer contextual details, the gist of original event, are retained in memory (Moscovitch, Cabeza, Winocur, & Nadel, 2016; Bartlett, 1932).

Besides memory stabilization, enhancement also occurs in memory consolidation particularly in association with sleep. During the enhancement phase of sleep-dependent memory consolidation, instead of memory decay, we can observe memory retention and we can automatically restore lost memories (Walker & Stickgold,2004).For example, memories of newly acquired words (Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010; Brown, Weighall, Henderson & Gaskell, 2012), temporal order (Griessenberger et al., 2012) and word associations (van Kesteren, Rijpkema, Ruiter, & Fernandez, 2010) improve after overnight sleep but not during the daytime. In addition, during memory consolidation the newly acquired knowledge would integrate with previous experiences, which may lead to more schematized memory. The schema-based consolidation was initially found in story reading. As reported in the early study by Bartlett, the memory for a story tends to fit into individual's schematic knowledge and own interpretation (Bartlett, 1932; Sulin and Dooling, 1974). Moreover, it also has been demonstrated that schema-congruent associations (e.g., classroom- chalk) is better remembered, relative to schema-incongruent associations (e.g., tennis court- soup ladle), indicated by the higher accuracy in recall and recognition tasks (Marlieke T R van Kesteren et al., 2013; Brewer and Treyen, 1981). Therefore, for representations of event duration, one might expect event schematic knowledge to be integrated with episodic memories during sleep, resulting in stronger influence of prior knowledge. In Chapter 3, we specifically test this predicition.

The next stage is memory retrieval, during which we usually seek for a particular memory (target memory). For example, how long was your breakfast this morning or what did you have for dinner yesterday?The factors determining retrieval success have been widely explored. It has been suggested that successful retrieval is affected by attention to cues, relevance of cues, cue-target associative strength, number of cues, strength of target memory and so on. In the studies of Chapter 3, I shall test whether the use of different type of retrieval cues (verbal cues VS. visual cues) would elicit different duration reproduction (Baddeley, et al., 2015).

1.2.1.2 Episodic memory and semantic memory

Semantic memory is the general knowledge about the world (Baddeley *et al*., 2009) which can be acquired through either real-world experiences or linguistic experiences such as reading and conversation. For example, when is Christmas? How long is a football match?

There is also another type of memory which is defined as ‘episodic memory’. Episodic memory is the memory of a specific event or a past episode which happened at a given time and in a particular place (Baddeley *et al*., 2009). For example, what did we have for our breakfast this morning? How long did it take to cook our dinner last night? Episodic memory allows us to re-experience or ‘re-live’ past events (Tulving, 1985). Moreover, the only way to construct an episodic memory about an event is to experience it personally in the real world.

There are many studies which have demonstrated the fundamental distinction between episodic memory and semantic memory, such as case studies of amnesia or brain damage (Kapur, 1999; Yasuda *et al*., 1997; Vargha-Khadem *et al*., 1997) and brain imaging studies (Wheeler *et al*., 1997). For example, amnesic patients with damage to the hippocampus are reported to lose their memory for specific events occurring either before or after the brain damage, but their language comprehension ability is not affected by the brain damage (Rempel-Clower *et al*., 1996; Nadel & Moscovitch, 1997; Stefanacci *et al*., 2000; Aggleton & Brown, 1999). These studies indicate functionally separate memory systems for episodic memory and semantic memory.

Even though episodic memory and semantic memory are demonstrated to be separate memory systems, they interact with each other during information processing. On the one hand, we extract semantic knowledge from multiple experiences. On the other hand, it has been suggested that semantic memory guides the encoding and retrieval of episodic memories. It has been found that participants with more background knowledge about a specific topic (such as baseball or chess) also have better episodic memory for information relating to the topic (such as the position of players in the field or pieces on a chessboard) because the knowledge stored in the semantic memory facilitates the encoding and retrieval (Ausubel, 1960; Chase & Simon, 1973). Whereas, there’s a long-lasting argument about when does semantic knowledge affects episodic memory at encoding stage or retrieval stage. We will discuss more about this later on in section 1.2.2.

1.2.2 Schema theory

One crucial part of semantic memory is schemas. Although there is no consensus about the definition of a schema, the earliest and most influential answer was given by Bartlett (1932) who suggested that schemas are integrated and organized pieces of knowledge about the world. For example, a general schema for ‘restaurant’ is composed of generalized information about situations, people and activities which occur in a place. Specifically, we assume that there should be tables, chairs and tableware in restaurants and that the main products of restaurants are professionally prepared food and drink. All the knowledge relating to the concept ‘restaurant’ is considered to be a schema.

Schemas have three fundamental characteristics, one of which is the inter-relationship between them. Schemas consist of a set of general frameworks which are linked together and which have an “associative network structure” (Ghosh & Gilboa, 2014). For example, the restaurant schema described above can be part of another schema such as ‘meal’ and might overlap with other schemas such as ‘café’ because they share common features. This feature reduces the total amount of information about this particular topic which is stored in the memory. It also enables us to digest acquired pieces of knowledge by relating them to each other.

Another characteristic of schemas is that they are extracted from memories of multiple episodes. Memories of multiple episodes are summarized in terms of their similarities and abstracted into the corresponding schemas as shown in Figure.1.1 (Bartlett, 1932). If schemas were based on a singular episode, they would lose generalization and be unable to guide people’s behaviour in a novel situation (Ghosh & Gilboa, 2014; Bartlett, 1932).

In addition, as schemas are composed of the commonalities of multiple episodes, they are characterized by the lack of details (Ghosh & Gilboa, 2014). This is because the details of episodes (the episodic memory of events) are forgotten gradually, so that we are only able to retain the general features of episodes. Moreover, the high-level and generalized nature of schemas enables their high flexibility, so schemas can efficiently and robustly accommodate different new situations. A schema may have variables to which we can assign distinct values. For example, when we stand in front of a restaurant, the schema of 'restaurant' is activated. Obviously, we know that there is an important variable, food. When we see the name of the restaurant, we know what type of restaurant it is. If it is an Italian restaurant, we assign 'Italian food', such as pizza or pasta, to the variable 'food'. But if it is an Asian restaurant, we assign 'Asian food' such as rice or sushi to the variable.

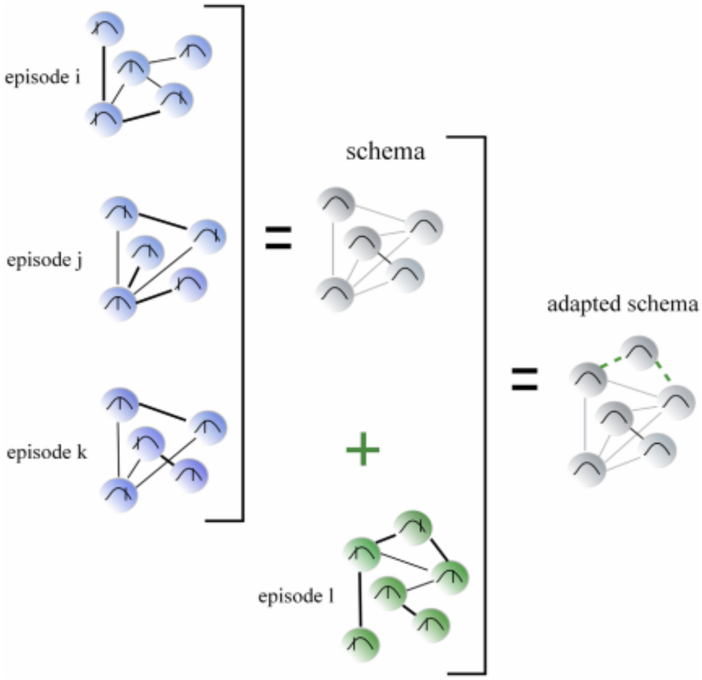


Figure 1.1 Schema theory: The small circles in one episode indicate the schema's units and the lines connecting the circles represent their interrelationships. 2) The blue and green frameworks (episode j, episode k, episode i) are the particular episodes and the gray networks are the schemas. 3) Multiple episodes are extracted into one schema based on their commonalities, and the existing schema is integrated with the new episode (episode l) to convert them all into the adapted schema.

Finally, schemas have the property of adaptability. Prior-existing schemas are not unchangeable but are always sensitive to novel information and adapt to new situations automatically. They are constantly assimilated and modified according to incoming new experiences. New episodes are either incorporated into the existing schema or used to revise the existing framework.

There is abundant evidence suggesting the essential role of schemas or schematic knowledge in information encoding and memory recall (for example, Alba & Hasher, 1983; Baddeley *et al*., 2009). On the one hand, knowledge of schemas enables us to predict upcoming information (Baddeley *et al*., 2009). When we process a novel situation, we construct expectations based on schemas. For example, O'Brien *et al*. (1998) found that reading schema-incongruent sentences leads to longer processing time. To be more specific, when reading the sentence 'Mary has been a strict vegetarian for ten years', readers would automatically make the inference that Mary will not order meat or cheese. But if the subsequent sentence is inconsistent with our inference or expectation (for example, 'Mary ordered a beef burger and fries'), our reading speed would significantly slow down. O'Brien *et al*. (1998) concluded that schematic knowledge enables us to construct expectations for upcoming information.

Similar findings were reported in an ERPs study (event-related brain potentials) carried out by Metusalem *et al*. (2012) who tested N400 during online language comprehension (N400 is the index of activation of semantic information) and found that in online language comprehension, there is reduced N400 when the target words are congruent with and related to the described events, indicating the activation of related schematic knowledge and integration with the current context (Altman & Mirkovic, 2009; Bransford & Johnson,1973; Kutas & Federmeier, 2011; Kutas & Hillyard,1980). The results showed that schematic knowledge is activated immediately in online reading and helps the reader to form expectations and assists language comprehension.

Schemas or schematic knowledge also influence memory recall. On the one hand, it can facilitate memory selectively. Specifically, schema-congruent information tends to be remembered more accurately whereas schema-incongruent information tends to be ignored and fails to be recalled from the memory. Palmer (1975) exposed participants to a picture of a scene (for example, a kitchen) and then presented them with a picture of either a schema-congruent object (such as a loaf) or a schema-incongruent object (such as a mailbox). The findings showed that memory of schema-congruent objects (the loaf) were more accurate than memory of schema-incongruent objects (the mailbox).

On the other hand, it has also been suggested that schemas lead to systematic distortion in the memory, which is called schema-driven distortion in memory. Bartlett (1932) investigated participants' errors in their memories of folk tales taken from an unfamiliar culture (American Indian) and extremely strange for those living in the western culture. Bartlet found that the participants tended to remember the story as shorter and more coherent. In addition, they tended to interpret the story in terms of their own views so they modified confusing details to fit into their expectations. Bartlett argued that the systematic errors in the participants’ memory were due to the interference with schematic knowledge. Following this idea, a number of better-controlled studies provided evidence in support of schema-driven errors in memory. For example, Sulin and Dooling (1974) discovered that the same stories with different main characters (Gerald Martin or Adolf Hitler) were remembered differently. Those who read the story about Adolf Hitler tended to believe that they had read sentences such as ‘He persecuted Jews’.

Although most memory research refers to schemas as pieces of knowledge, language is the most common way to convey schematic pieces of information, as indeed, word meanings are just this sort of information. Bartlett’s research (1932) for example used different story titles to manipulate the prior knowledge that readers would bring to mind in reading the stories. However, when it comes to language influences on memory, there is has been an enduring debate on whether and how schematic knowledge conveyed by language shapes episodic memories. In general, there are two views: the interactive encoding accounts and the interactive retrieval account. The interactive retrieval accounts have argued that language labels bias memory at retrieval stage, as they provide retrieval strategies to recollect episodic details (Alba & Hasher, 1983). For example, language labels presented at encoding stage may bias the reproduction of ambiguous drawings (Carmichael, Hogan, & Walter, 1932). However, the presence of labels at encoding would not influence the recognition of the ambiguous drawings. The finding suggested that the initially encoded representation was not distorted by the labels and the language-induced distortion took place at retrieval stage (Carmichael, Hogan, & Walter, 1932; Hanawalt, 1937; Prentice, 1954; Rock & Engelstein, 1959). Moreover, the language-induced memory distortion can occur even when linguistic concepts were absent at encoding stage (Hanawalt & Demarest, 1939). For instance, it has been found that misleading questions (e.g., how long did it take the man to walk/run through the lecture theater?) presented to witnesses one or two weeks after the exposure to the original events (e.g., a man burst into a lecture theater) would distort their numerical estimations of event duration or object speed in the original event (e.g., *walk* vs *run*, *collide* vs smash *;* Burt & Popple, 1996; E. F. Loftus & Palmer, 1974). These studies highlighted the potential language influences at retrieval stage during information recollection.

On the other hand, the interactive encoding account proposed that: the exposure to language labels at encoding would activate conceptual features that distort perceptual encoding via top-down feedback (Lupyan, 2008, 2012). This process leads to the encoding of distorted visual stimuli, resulting in failures to recognize previously learned stimuli (Pezdek et al., 1986, Lupyan, 2008, Feist & Gentner, 2007). Nevertheless other studies examining concurrent verbal and visual event encoding did not show language effects on memory (Gennari, et.al, 2002; Papafragou, Hulbert, & Trueswell, 2008; Trueswell & Papafragou, 2010). Hence, the underlying mechanism of language influences on memory remains unclear. To solve this question, in Chapter 3, we will carry out a series of experiments and manipulate the presence of language label (at encoding stage or retrieval stage) to explore whether and how language labels shape memory for event durations.

Apart from the unsolved issues regarding the role of language, schema theory is useful in highlighting the key role of our prior experiences or world knowledge in processing new experiences. In reading, we used world knowledge to create mental representations or models of the events being described, as discussed later below. In event perception, prior knowledge is used to guide the way the information is segmented and the predictions we make as event unfold. Thus, schemas and stored knowledge more generally, are critical in understanding the world around us, and for this reason, they play an important role in most theories of event representations and language comprehension. We turn to these theories below.

1.2.3 Event segmentation

We live in a stream of ongoing, dynamic, multimodal information, but we perceive the world as a series of discrete events (Newtson, 1973; Zacks & Tversky, 2001; Zacks, 2004). To make sense of this continuous information, we use our existing knowledge to identify salient boundaries and segment the stream into meaningful units, or events, which is known as event structure perception (Zacks & Tversky, 2001; Radvansky & Zacks, 2011; Zacks, 2004). Event structure perception is a basic cognitive mechanism which plays a key role in our ability to perceive, comprehend and predict upcoming information and also contributes to memory recall (Abelson, 1981; Grafman, 1995; Zwaan & Radvansky, 1998).

1.2.3.1 Event segmentation theory

EST was first proposed by Zacks *et al*. (2007). The input of the proposed system is continuous sensory information which is then passed to the perceptual predictions as the output (*see* Figure 1.2). The perceptual predictions are rich in semantic content, including information such as agents, goals, location and motion trajectories. It has been argued that perceptual predictions are crucial because they enable us to predict the future and plan actions; for example, by successfully predicting the future positions of moving objects, or by comprehending others' goals to predict their future actions.

When people process ongoing perceptual information, they are guided by event models to know what is happening now. An event model offers a stable representation of an ongoing event. Hence, people can process the perceptual input at a specific level of constancy. An event model is constructed from both bottom-up and top-down information. On the one hand, when it constructs a new event model, it receives perceptual and sensory input from multiple modalities (for example, visual and auditory). Additionally, an event model is also driven by an event schema which enables people to relate upcoming information to prior experiences according to their common features. People can then make predictions about upcoming perceptual changes.

A crucial part of EST is an error detection mechanism which is implemented to monitor the quality of an event model. It is used to detect how predictive the current event model is so that it can be updated if necessary. For example, when there is an increase in errors in perceptual predictions, a new event model is required. In this condition, the gate linking the event model and sensory input is activated and they are connected so that the event model can receive the perceptual input and be updated to a new event model. When the perceptual predictions are relatively reliable and there is no need to change the event model, the path between sensory input and event model is disconnected. Critically, the periods of updating are considered as event boundaries, and the periods of stable representations are perceived as events.

To explain clearly how EST works, Zack *et al*. (2007) gave a good example. Think about a daily situation, for example a man is washing dishes. The man takes a dish from a pile, wipes it using a sponge and then places the plate in the sink. The same actions are applied to the second dish. At that moment, because we infer that the man will continue to clean all the dishes in the pile, we expect to see him continuously doing the same to more dishes. His movement patterns are therefore predictable. When he finishes the final dish, his actions become less predictable because the previous movement pattern cannot apply to the current situation. At this point, therefore, there will be increased errors in the event model so the gating mechanism will be activated and the event model will be updated.

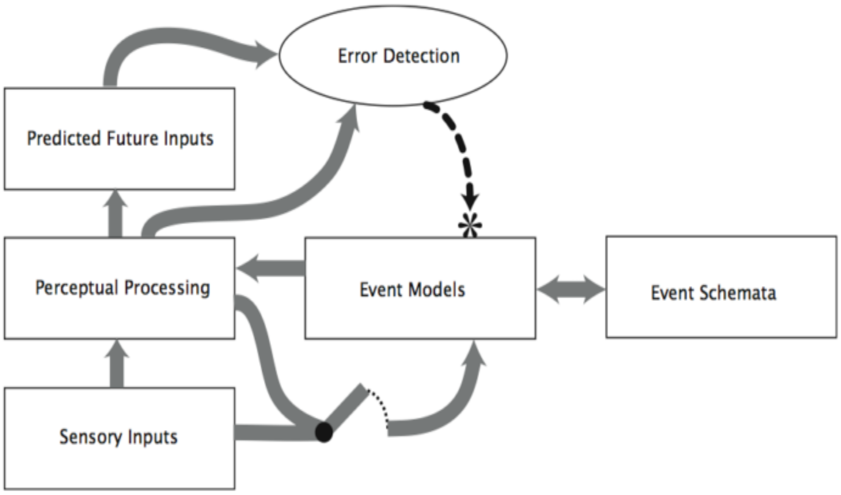


Figure 1.2 Event segmentation theory: a diagram describing EST (adapted from Zacks et al., 2007: 2). 1) The grey arrows connecting the processing areas represent the information flow and direction between them. Sensory input flows to perceptual processing, where the event model is implemented to guide the perceptual processing and make perceptual predictions. 2) The dashed line between error detection and event models shows the error detection mechanism which is used to evaluate the quality of perceptual predictions. 3) The gate between event models and sensory input is only activated in periods of updating or resetting. 4) Event models also receive information from event schemas, including prior experience of similar events.

There is abundant evidence in support of the event segmentation model. It has been reported that participants were remarkably consistent when identifying events in a short film: they typically agreed about when one event finished and another started, suggesting that similar criteria were used to identify events across people. (Newtson, 1976; Speer *et al*., 2003; Zacks, Speers et al., 2009). Similarly, evidence from reading studies shows that there is decreased reading speed at event boundaries (for example, changes in causality, characters, space or time) (Zwaan, Magliano & Graesser, 1995; Zwaan, Langston & Graesser, 1995; Zwaan & Radvansky, 1998). For example, Speer and Zack (2005) manipulated the temporal contiguity of narratives by using different temporal phrases and had participants read narratives of either temporal contiguity (such as ‘a moment later’) or discontinuity (such as ‘an hour later’). It was found that reading time increased at phrases marking temporal changes (temporal discontinuity). They argued that event boundaries lead to processing costs, as memory updating mechanisms take place at event boundaries: storing the previous events to a longer-memory and starting a new event model (and new predictions). Despite the behavioural evidence, brain image studies have also supported the automaticity of event boundaries. Zacks *et al*. (2007) found that when participants were passively viewing a film, their brain activity increased significantly just before marked event boundaries and peaked at event boundaries. These results suggest that event segmentation is an automatic process and that we tend to automatically or spontaneously segment upcoming information into distinct events and that segmentation is associated with extra processing costs leading to longer reading time and brain activity.

1.2.3.2 Hierarchical structure of event segmentation

It has been suggested that events can be segmented as a range of temporal grains, with coarse segments composed of multiple lower-level, fine-grained segments (Zacks & Tversky, 2001). For example, the events involved in ‘cooking potatoes’ can be decomposed into sub-events such as preparing the potatoes, frying or boiling them and then possibly placing cooked potatoes in some container. In addition, preparing the ingredients can be further divided into ‘peeling potatoes’, ‘slicing potatoes’ and so on. It has been found that adults’ viewing time of action sequences increases significantly at coarse boundaries

(Hard *et al.*, 2011). This hierarchical organization allows viewers to chunk information at different levels and enables the storage of more complicated information (Ghosh & Gilboa, 2014). Here the question arises of how viewers extract the hierarchical structure from the continuous ongoing information. This process is not yet fully understood. In general, there are two explanations for hierarchical segmentation: the physical cue-based explanation and the conceptual cue-based explanation.

A physical basis to hierarchical event segmentation suggests that bottom-up perceptual changes are sufficient to cue the hierarchical organization of events. It has been shown that the hierarchical organization of an event can be detected in the absence of knowledge of the goal-hierarchy. It has been found that infants and animals such as primates and rats, which have no knowledge of goal-hierarchy and event structure, can nevertheless detect the hierarchical structure of an event (Baldwin *et al.*, 2001; Woodward & Sommerville, 2000; Woodward, 1998). For example, Baldwin *et al.* (2001) found that infants aged ten to eleven months were sensitive to goal-completion and able to parse dynamic everyday actions with respect to goal completion. It was found that a pause in an action sequence occurring before the completion of the goal elicited greater surprise and increased interest in video recordings compared with when the motion paused after goal completion. They explained that physical changes which varied in quantity or quality usually coincided with the completion of goals at multiple levels (Hard *et al.*, 2006; Zacks *et al.*, 2001; Zacks, 2004). So the perception of physical breaks enables us to achieve hierarchical segmentation, independent of awareness of intentions or event structures.

By contrast, instead of highlighting the influences of physical cues on predicting hierarchical segmentation, the conceptual cue-based account emphasizes the top-down influence from event schemas, especially the knowledge of goal/sub-goal relationships (Zacks & Tversky, 2001). It has been suggested that the identification of fine-grained segments is based on movement features whereas clustering the fine-grained segments into larger units is primarily directed top-down by knowledge of the goal/sub-goal hierarchy (Zacks, 2004). For example, it has been found that knowledge of event's aim can elicit more hierarchical segmentation (Newtson, 1973; Zacks *et al*., 2001). Hard *et al.* (2006) found that exposure to linguistic concepts (for example, hide and seek) resulted in fewer segments for abstract movies. The explanation for this was that the linguistic concepts activate the goal-hierarchy stored in semantic memory and the fine-grained units can therefore be grouped into larger chunks. In addition, people tend to parse events in terms of the most superior goals that they can recognize (Wilde, 1978a, 1978b), so conceptualization will elicit coarser segmentation. However, the conceptualization effects in event segmentation are inconsistent. A similar study by Zacks (2004) failed to observe the coarser segmentation as a function of conceptualization. In that study, participants were asked to segment abstract animations of geometric figures which were described as moving randomly or intentionally (for example, playing a video game). The findings did not replicate the tendency for coarser segmentation of the conceptualized videos.

To sum up, the roles of top-down conceptual knowledge and bottom-up perceptual information in cueing hierarchical segmentation remain unclear now. Moreover, the conceptualization effects in event segmentation are inconsistent across studies. The precise interpay between schemas and bottom-up information is an important issue to elucidate because event segmentation is known to determine event memory, i.e., what people later remember about events is linked to how they initially segmented the stimuli. This is discussed below. Thus, event memory and event segmentation are intimately related. To further understand the role of schmas on event segmentation and memory, in Chapter 5, I shall further explore the factors leading to the inconsistent conceptualization effects in previous studies and reveal the contributions of schemas in cueing event segmentation.

1.2.3.3 Event segmentation and memory

Segmentation affects memory because event boundaries serve as the anchors on which viewers can organize encoded information (Loucks & Meltzoff, 2013; Jeunehomme *et al*., 2018; Zacks & Tversky, 2001). According to EST, perceptual information which is processed at event boundaries should be remembered better than information not at a boundary. For example, Newtson and Engquist (1976) explored how event boundaries affect the recognition memory of videos of everyday activities such as looking for something or cutting cloth. The participants watched videos first and then identified whether a given still frame was from the video which they had watched. It was found that accuracy of recognition was higher for the frames appearing at event boundaries than for those at non-boundaries. This result supports EST in that at event boundaries, to update the existing event model, more extensive perceptual information is processed, which leads to better memory for details appearing at event boundaries (Zacks *et al*., 2007).

Additionally, events encoded at different grains are remembered differently. Compared with coarser-grained segmentation, fine-grained segmentation leads to more recalled details. Hanson and Hirst (1989) asked participants to segment events in a video into fine-grained or coarse-grained and then describe the video from their memories. It was discovered that those participants who had been required to segment events into fine-grained segments provided more details in recall tasks. It was concluded that richness of memory is influenced by the grain of segmentation. Boltz (1992) found that videos with commercial breaks at event boundaries were remembered just as well as those without any commercial breaks. In contrast, when commercial breaks appeared between boundaries, the memory of the videos was impaired (Boltz, 1992; Schwan *et al.*, 2000; Schwan & Garsoffky, 2004). Consistent with these findings, it has also been observed that the absence of information at event boundaries leads to impairments in event processing and memory recall (Schwan & Garsoffky, 2004). Taken together, these results suggest a critical role of event boundaries in organizing memory, even when people are not explicitly segmenting.

Event segmentation is also regarded as a specific ability and people with better segmentation ability have better memory performance (Zacks *et al.*, 2006; Sarget *et al.*, 2013). In those studies, participants were required to segment movies of everyday events into meaningful units by pressing a button. Segmentation ability was considered to be the degree of agreement between individuals' segmentation and the whole sample's segmentation at event boundaries. It was found that participants with better segmentation abilities also had better memories of the movies. In contrast, older adults, especially those with mild dementia, were found to have weaker segmentation ability compared with younger adults, which possibly explains the memory deficit in old people (Zacks *et al*., 2006).

Critically to our present focus on duration representations, event segmentation is also suggested to affect remembered durations of past events. For example, events with more segments are remembered as longer (Avni-Babad & Ritov, 2003; Bailey & Areni, 2006; Block, 1978; Boltz, 1995, 2005; Faber & Gennari, 2015; Jeunehomme *et al.*, 2017; Poynter, 1983, 1989; Zakay *et al.*, 1994), and events with hierarchical structures are remembered as shorter. We will review more about this later on in section 1.3.

Taken together, the evidence in support of EST suggests that 1) event segmentation is automatic; it occurs spontaneously and automatically when people experience a continuous information stream; 2) event segmentation is determined by perceptual cues and prior knowledge; by combining information on sensory input with their prior knowledge, people construct an event model representing ongoing experience and can predict upcoming information; However， the specific role of bottom-up conceptual information and top-down perceptual information in cueing event segmentation remain unclear and I shall further explore the question in Chapter 5; and 3) event segmentation plays a crucial role in event perception and memory, and the way that an event is segmented determines the way it is remembered. It has also been suggested that event segmentation affects the memory of time, and this will be discussed in more detail later in this chapter.

1.2.4 Situation model

1.2.4.1 Situation model theory

As has been explained above, according to EST, people automatically segment an ongoing dynamic information stream into meaningful events. However, how do we process and comprehend an event in terms of language comprehension? According to the situation model, it has been proposed that when we process events in language, instead of encoding individual words as texts, we construct a mental model or a situation model of the described events (Van Dijk & Kintsch, 1983; Rinck & Bower, 2000; Rinck & Weber, 2003; Zwaan & Radvansky, 1998; Zwaan, 2008, 1996; Zwaan *et al.*, 2000).

Zwaan and Radvansky (1998) suggested a detailed model of how people construct mental representations during language comprehension. They proposed the following three stages in constructing situation models: (i) first, when we initially encounter a new situation, we create an initial situation referring to time, space, causation, intentionality and protagonist(s); (ii) after that, with the unfolding of upcoming sensory information, we constantly integrate new information into the current model which becomes an integrated model; (iii) finally, after reading the whole text, we store internal representations into our long-term memory, and this is called a completed model.

Situation model has several key features. First, situation models are multi-dimensional because they are constructed on the basis of five aspects: time, space, causation, intentionality and protagonist(s) (Rinck & Weber, 2003; Zwaan, 2008, 1996; Zwaan *et al.*, 2000; Zwaan & Radvansky, 1998). Additionally, situation models are regarded as being dynamic and evolving. During language processing, we constantly keep track of information in the five relevant dimensions. When we first read about an event, we index it under these five aspects. As we read about subsequent events, we identify whether any of the five indexes should be updated. If they overlap into more indexes with the previous event, the incoming events can be more easily integrated into the situation model. If not, a new situation model will be constructed which will lead to a larger processing cost (Zwaan & Radvansky, 1998; Kelter *et al.*, 2004). For instance, reading stories containing inconsistent emotions would slow down the reading speed compared with reading stories containing consistent emotions because the emotion shift would elicit a processing cost (Vega, 1996). Similarly, if comprehenders read the sentence ‘Mary is a vegetarian’ followed by an incongruent action such as ‘… ordered a beefburger’, there will be an increase in processing time because of the difficulty in integrating an opposite protagonist’s traits (O’Brien *et al*., 1998).

According to the situation model theory, internal representations of language are grounded in our world knowledge and sensorimotor perception. When one event is encountered in language, we automatically re-enact the perception referring to the described events and activate its associated world knowledge. Many studies provided evidence in support of mental simulations in situation representations (Zwaan & Pecher, 2012; Stanfield & Zwaan, 2001; Palmer, 1999). For instance, Palmer (1999) found that exposure to sentences involving an object (for example, clothes) would prime a picture of the described object. For example, reading the sentence ‘the clothes are in the drawer’ would elicit faster reaction to a picture of folded clothes than to unfolded clothes, whereas exposure to the sentence ‘the clothes are on the hangers’ would prime a picture of 'unfolded clothes' rather than a picture of 'folded clothes'. This is explained as that when we read the sentences, even though there is no direct description of the state of the clothes, by making use of our world knowledge we visually simulate the situations described.

1.2.4.2 Time in the situation model

Time is a crucial part of the situation model (Rinck & Weber, 2003; Zwaan, 2008; Zwaan, 1996; Zwaan *et al.*, 2000) and readers always track temporal information, which influences on-line comprehension and later retrieval (Rinck & Bower, 2000; Vigliocco *et al.*, 2004; Kelter *et al.*, 2004; Gennari, 2004; Coll-Florit & Gennari, 2011; Zwaan, 2008; Carreiras *et al.*, 1997; Zwaan, 1996). Zwaan (1996) suggested that the duration of time shifts in narratives leads to the construction of new situation models or the updating of current situation models; he compared stories containing a short time shift (for example, a moment later) with stories containing an intermediate time shift (an hour later) or a long time shift (a day later). He found that relative to a short time shift, a longer time shift led to shorter reading times and response times. These results provide evidence supporting event-indexing models in that temporal information is involved in mental representations. They also suggest that discontinuity in the temporal dimension (for example, an hour or a day) will increase the difficulty of incorporating upcoming information into the current model. Hence, a new situation model should be constructed. Consequently, the reading time for the subsequent event is longer and the previous event is less accessible.

In line with those findings, Gennari (2004) found that reading times were shorter when the temporal references in the subordinate clauses were closer to the time of the main clause events. For example, a distant temporal reference is ‘The minister admitted two days ago that diplomatic relations with Cuba were difficult this year’, and a close temporal reference is ‘The minister admitted two days ago that diplomatic relations with Cuba were difficult last month’. The findings also suggest that temporal information has an impact on the cost of updating current models.

Even though an extensive body of research has investigated the impact of time shifts in narratives, few studies have focused on representations of event duration. Kelter *et al.* (2004) provided evidence demonstrating that longer events decrease the accessibility of previously presented events. In that study, participants read stories consisting of two consecutive events. The first sentence in the story was identical in both the short and the long version, but the second sentence indicated either a short duration or a long duration (‘bake some cookies’ as opposed to ‘put some cookies’). They were then asked to retrieve information about the first sentence. It was found that the first sentence was less accessible when the second sentence referred to longer events. In short, the participants needed more time to retrieve a probe (a word appearing in the first story) in stories with long events (such as ‘bake some cookies’) compared with stories with short events (‘put some cookies’). The researchers suggested that the longer the temporal distance between two events, the less accessibility there will be to the previously presented event. These findings provide evidence that duration information is encoded in the mental model. However, instead of focusing on the representation of event duration itself, the findings emphasized how the temporal relationships between two events influence event representation. In a related study, Coll-Florit and Gennari (2011) looked specifically at the representations of event duration in on-line language comprehension. They examined how people process event duration which is conveyed by verbs. They discovered that people take longer to process durative events (for example, ‘owe five euros’) than non-durative events (‘lose five euros’) and suggested that longer events require longer to process because they relate to more diverse background knowledge.

In summary, the studies discussed above suggest that temporal information is involved in the mental representation of described events and influences language comprehension and memory. Furthermore, the situation model explains the results as being caused by the updating mechanism of the situation model. However, the studies focused primarily on the role of temporal information in cueing model updating, that is, how the temporal relationships between successive events affect event representations; but the representations of event duration of singular events have been largely ignored, except by Coll-Florit and Gennari (2011). However, their study manipulated event duration by using verbs indicating long or short duration. Therefore, the observed duration effects were possibly due to the activation of different event schemas (verbs) rather than the differences in event durations. Moreover, duration information has to be inferred rather than being explicitly presented in the stories. Explicit mention of event durations in a narrative would thus be necessary to better control for the verb effects and further explore explicit duration representation of a singular event in narrative. Hence, in this thesis, therefore, in Chapter 2, I shall manipulate event duration via temporal adverbials to explore explicit duration representation of a singular event.

1.2.4 Embodied Theory

Embodied theory is also used to explain representations of events. It was suggested that by mentally simulating the described events, people had access to the perceptual representation (such as shape, size) and the actions it affords (drinking, grabbing, pouring), so that readers can understand the events. A large body of studies has provided evidence supporting this theory. Hauk, Johnsrude and Pulvermüller (2004) found that processing action words referring to face, arm or leg actions in a reading task differentially activated areas that were overlapped with areas activated by actions of tongue, fingers or feet. The results showed that people comprehend events by mentally re-enacting the described events.

However, it should be noted that time is an invisible, abstract concept. Then how does the embodied theory explain the representation of event duration? According to embodied theory, it has been suggested that we process time by relying on other factors such as the speed of movements or the distances between two locations, as longer duration is typically related to longer distance or slower movement. Therefore, the simulation of longer events with different duration can be considered as the simulation of slower movements or movements between two more distant locations (Lindsay, Scheepers & Kamide, 2013; Matlock, 2004).

For example, in the study of Matlock (2004), participants were exposed to narratives describing a trip of either long journey or a short journey. After reading the story, the participants were asked to retrieve information relating to the story. Critically, they manipulated the durations of the trip by changing the distance (short VS long), the difficulty (easy VS difficult) and speed (fast VS slow) of the journey. In addition, the target sentences were depicted by either a fictive motion (the e.g, road runs through the valley) or a non-fictive motion (e.g., “the road is in the valley”). The results showed that participants responded faster for the stories of the short, easy and fast-speed journey, only when journeys were described by fiction motion, whereas such effects were not observed among stories with non-fictive motion. The findings highlight the importance of verbs in eliciting simulation. Moreover, they suggested that as the longer distance, more difficulty and slow-speed usually implicated a more time-consuming trip, people take longer to simulate the events and took longer to retrieve information relating to the stories.

Similarly, in the study of Fecica & O’Neill (2010), they explored embodied theory among preliterate children and tested how expected movement duration (e.g., “walking” VS. “driving”) influenced language processing. They found that processing times were longer, when children read storeis describing events indicating longer durations (e.g., walking to the park) than when they read stories describing short events (e.g., driving to the park). They explained that as children simulated the shower action depicted by the story, they took longer to mentally simulate of the events. Moreover, Lindasy, Scheepers and Kamide (2013) also carried out an eye tracking study and they found evidence in support of the embodied view. In their study, they used visual world paradigm to test how participants matched visually presented scenes to auditory sentences. Specifically, they constructed sentences of either faster version or slow version by using different verbs (e.g., to dash VS to dawdle). In addition, while participants hearing sentences, they were guided to view a picture of the described scene with a path leading to the target objects. It was observed that participants launched more and longer eye movements to the path when they were exposed to sentences of slow movements (e.g., to dawdle) relative to when they were shown the sentences of fast movements.

However, the studies reviewed above did not strictly investigate the effects of event duration in language processing. Specifically, duration representations in the studies were all grounded in verbs or motions. They manipulated the speed or difficulty to construct longer and shorter events. Hence, it's not the explicit duration representation. However, in our daily lives we also communicate duration information explicitly by using temporal adverbials (e.g., It took me three months to read the book). Then, how do we represent the temporal information not grounded in actions/ movement?

According to the embodied theory, it would be predicted that longer events would take longer to process (e.g., playing basketball for half an hour VS. Playing basketball three hours), as people should simulate the temporal development of the described events time. Therefore, the simulated event duration should be analogous to the described event duration. However, in an obvious sense, simulation of events should not be the exact replica of real events. People simulate a depicted event in a temporally compressed manner. To be specific, to understand the sentence--Tom played basketball for two hours, we do not take two hours. However, the embodied theory did not come up with a clear explanation to the mechanisms. Therefore, alternative explanation is needed to explain explicit duration representation in language. To solve this question, in Chapter 2, we would focus on the explicit duration representation in language by manipulating temporal adverbials. In addition, we would also explore the underlying mechanisms of explicit duration representation in language and would suggest a new explanation for duration representation.

Taken together, situation schemas, event segmentation, schema theory and embodied theory focus on different aspects of event representations. Schema theory concentrates principally on static organized world knowledge, event segmentation suggests how people parse an ongoing dynamic information stream into discrete events, and the situation model focuses primarily on mental representations of events in language comprehension. Embodied theory proposed that people comprehend events by mental simulation. Despite these differences, schema theory, event segmentation theory, the situation model and embodied theory all emphasize the importance of schematic knowledge or prior event knowledge in guiding encoding, memory recall and language comprehension. Moreover, language has the power to activate the related schematic knowledge and then shape our perception and memory for events. These four theories provide the theoretical basis for understanding the representations of events and event duration.

1.3 Processing duration from real experience

This thesis was designed to explore how human beings process event duration in language and in memory. In the previous section, I reviewed the theories and literature on how we process events in general (schema theory, event segmentation theory and the situation model) and it can be concluded that event representation is guided by our prior knowledge. In order to explore the temporal dimensions of an event further, I shall now present and discuss the theories and experimental evidence about the representation of time in real-world experiences.

In our everyday lives, we continuously experience events and are aware of time passing. It is crucial for us to keep track of time and to be aware of event duration. For example, in witness testimony, temporal information about an event usually contains abundant information. Eyewitnesses are asked questions relating to event duration. For example, ‘How long was it before the man ran out of the bank?’ In our everyday lives, however, we do not always keep track of clock time using watches or cell phones. Instead, for most of the time, we rely on our sense of time and retrieve time from our memory. For example, when we are waiting for a traffic light to change, we can roughly estimate when the light will turn green without checking a watch. We can also easily answer a question about how long ago we had breakfast this morning. How do we manage to do this? What are the underlying mechanisms? Are there differences between timing an interval prospectively and recalling it from memory? Does our experience or world knowledge make a difference to our perception and memory of time?

In this section, I shall address these questions by reviewing the literature on the temporal aspect of events. First, I shall look at two different measures of time experience: the prospective paradigm and the retrospective paradigm. Specifically, how does timing an interval differ from reconstructing a duration from memory? Second, I shall consider the basic theories on prospective timing and remembered time. Finally, I shall discuss how our world knowledge or prior experience influences our timing and memory of duration. This section thus provides the background knowledge on duration representation from real-world experience and provides a better understanding of the questions that I am about to address in Chapter 3 & Chapter 4.

1.3.1 The prospective and retrospective models

In psychological studies, there are generally two types of event duration considered: experienced event duration and remembered event duration (Block, 1989). Experienced event duration is more about timing an ongoing event, but the remembered event duration is about retrieving duration information of a past event from memory (Block, 1989). It has been demonstrated that the differences between remembered duration and experienced duration are clear (Hicks *et al*. 1976; 6James 1890; Block, 1989). In the thesis, in Chapter 3 and Chapter 4, my focus is primarily on remembered event duration. However, to have a better understanding of remembered duration, it is important to know how the two types event duration differ.

James (1890) was the first researcher to point out the distinction between prospective and retrospective duration experiences. He suggested that prospective time experience (experienced time) is based on cognitive timing, whereas retrospective time experience (remembered time) is reconstructed from memory. In psychological experiments, the two types of time experience are tested by prospective paradigms and retrospective paradigms respectively. In prospective paradigm experiments, the participants are aware of the timing tasks in advance, whereas in retrospective tasks, they have no idea of a timing task when they are exposed to the stimulus, but later they are asked to judge the duration of the stimulus (Hoerl & McCormack, 2001). It should be explained that in the prospective paradigm, because the participants are informed of the timing task in advance, they will intentionally make an effort to process the temporal aspect of the stimulus (Hoerl & McCormack, 2001), but in the retrospective paradigm, the participants are ignorant of the duration judgment task when processing stimuli; their judgment is based on their incidental encoding of temporal information and they reconstruct duration from relevant information stored in the memory (Block & Zacky,1996). It is therefore suggested that the prospective paradigm tests ‘experienced duration’, whereas the retrospective paradigm examines ‘remembered duration’ (Zakay & Block, 1997).

Investigations exploring the distinction between prospective and retrospective duration judgment were not common until Hicks *et al*. (1976) required participants to sort playing cards into one stack, two stacks (according to colour) or four stacks (according to colour and suit). The participants should process less information if they are required to sort the cards into fewer stacks. They were also either informed (the prospective paradigm) or not informed (the retrospective paradigm) about the timing task before sorting the cards. It was discovered that in the prospective paradigm, participants estimated duration as shorter when they sorted the cards into more stacks, whereas in the retrospective paradigm, there was no correlation between the number of stacks and duration judgment. The researchers explained that in the retrospective paradigm, because the participants did not intentionally memorize their activities during the task, they did not have any reference to the amount of processed information, whereas in the prospective paradigm, sorting the cards into more stacks would reduce the attention assigned to the temporal aspect, which decreased the estimated duration. Block (1992) later replicated the study and obtained similar results. Specifically, in the prospective paradigm, experienced durations were lengthened when the participants did the easier processing task, but the effects were not observed in the retrospective paradigm. These findings support the idea that prospective timing differs from retrospective timing.

More supporting evidence was provided by McClain (1983) and Predebon (1996) and the meta-analysis carried out by Block *et al.* (2010). McClain (1983) explored the relation between duration judgment and processing level in the prospective paradigm. Words were used as the stimulus and participants were asked to process either the graphemic form (shallow-level processing) or the semantic meaning (deep-level processing) of the words. The results showed that in the prospective paradigm, duration estimation was shorter in deep-level processing than in shallow-level processing, as processing semantic meaning leads to higher processing demands. In contrast, in the retrospective paradigm, duration judgments were related positively to the amount of remembered information (length of the word list),but were not affected by processing level.

Predebon (1996) examined how the correlation between time judgment and stimulus quantity varies as a function of processing conditions: active processing (responses are required) and passive processing (responses are not required). The results showed that in retrospective time judgment, duration judgments varied as a function of the number of stimuli, regardless of processing conditions (passive or active processing). In the prospective paradigm, however, the estimated durations were lengthened by an increased stimulus quantity in the active processing condition.

In general, the differences between remembered duration and experienced duration are clear (Block, 1989). It has been suggested that experienced duration and remembered duration have been demonstrated to be different as they rely on different processing systems (Hoerl & McCormack, 2001). Specifically, in the retrospective paradigm, participants have to reconstruct duration on the basis of their memory (Hoerl & McCormack, 2001). In the prospective paradigm, however, duration estimation is largely determined by the attention allocated to temporal information during the time period and cognitive load of stimulus and tasks (Block *et al.*, 2010). In this thesis, in Chapter 3 and Chapter 4, the focus is primarily on remembered event duration, but it is important to know the differences between remembered experienced duration.

1.3.2 Models of human timing

In the previous section, I discussed the differences between experienced time and remembered time. In this section, I shall consider the models explaining time representations. They can be roughly categorized into clock-based models and cognitive models. Specifically, the internal clock model focuses on prospective duration judgment whereas the cognitive model emphasizes the role of memory in duration judgment and focuses on retrospective time. Because I am focusing on retrospective time in the current study, I shall place emphasis on the cognitive models explaining retrospective time estimation and give only a brief overview of timing models.

1.3.2.1 Internal clocks and representations of duration

The main idea of internal clock models is that animals and human beings perceive time by means of an internal clock. The core notion is that humans and animals possess an internal clock, also known as the pacemaker-accumulator system. Specifically, a pacemaker produces pulses which flow to an accumulator where the number of pulses is counted. When more pulses are flowing to the accumulator, we perceive a longer duration (Hoerl & McCormack, 2001: 38; Grondin, 2001). One of the more recent clock-based models was proposed by Gibbon *et al.* (1984): the scalar expectancy theory (SET). This model suggests a constantly working pacemaker which produces pulses at a random rate (*see* Figure 1.3). In addition, there is a switch connecting the pacemaker and the accumulator, and this switch is only activated when timing a stimulus. The pulses can then flow to the accumulator. When the timing ends, the number of pulses collected by the accumulator serves as the basis of duration judgment and is then stored in the working memory or the long-term memory (as described by Matthews *et al.*, 2011). The final part of the model is a comparison mechanism or duration decision process. The comparator calculates the value of an individual stimulus and compares it with the value of a reference stimulus in the memory. Moreover, SET also suggests that the timing, storage and comparison processes might be influenced by changes such as stimulus intensity (Zelkind, 1973) and switch latencies (Wearden *et al.*, 1998; Matthews *et al*., 2011). When we perform timing tasks, the three processes (clock, memory and decision) work together and lead to the observed timing behaviour.

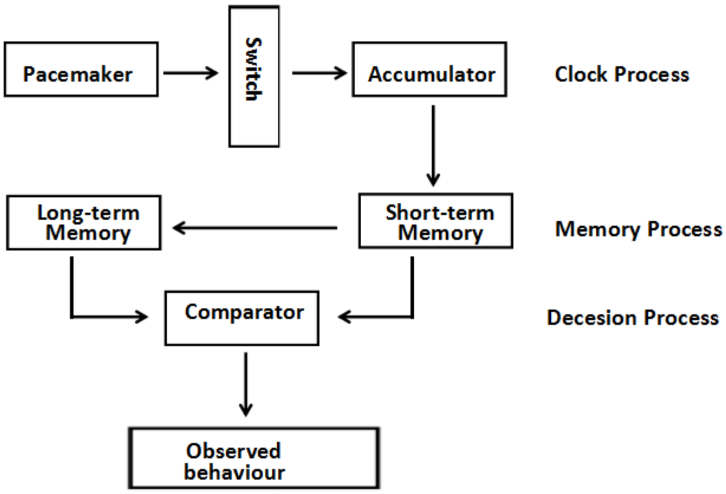


Figure 1.3 The internal clock model: the outline of the information processing model of SET. The first part shows the Clock Process containing a pacemaker, an accumulator and a switch connecting them. The second part is the Memory Process, here the contents of the accumulator are stored in the short-term memory and then some of the contents are transferred to the long-term memory. The final part is the Decision Process comprising a comparison between the value of an individual stimulus with a reference value. The results of the Decision Process lead to the observed behaviour.

One of the most popular paradigms of the internal clock model was introduced by Wearden (1991) and was based on a comparison between current duration and a standard duration. Specifically, in timing tasks, participants were familiarized with a standard duration and then had to compare a new stimulus with the standard duration and identify whether the two intervals were the same in length. In addition, Wearden (1991) suggested a scalar representation of duration, a linear relationship between clock time and psychological duration.

There are limitations to internal clock models. First, it has been suggested that different time ranges, below one second or above one second, seem to require different processing mechanisms (Penney & Vaitilingam, 2008; Gibbon *et al*., 1997). Short intervals below one second are based on automatic processing but experience of longer durations can hardly be explained by clock-based theory unless memory updating mechanisms are proposed (that is, how we chunk and store information beyond the limitations of our working memory). Moreover, the internal clock model cannot explain the results of attention-demanding tasks or dual tasks and it has been suggested that people's timing performance will be interrupted by cognitive mechanisms or resources (Brown, 2008). Ultimately, performance in timing tasks depends on how much attention is allocated to accumulating pulses, as opposed to other concurrent tasks. Finally, how we represent time from memory, for example when thinking back to past experiences, does not seem to be captured by these timing mechanisms which are meant to explain online timing processes.

1.3.2.2 Cognitive models of duration memory

In contrast to the internal clock model, other researchers have suggested the cognitive-based model, which explains the representations of time from the perspective of memory, segmentation and changes. Two cognitive models will be reviewed and discussed in this section: the storage size model and the change-based model. The storage size model proposes a correlation between mental storage size and remembered duration, whereas the change-based model explores how the number of changes in past experience relate to subjective duration. As a third alternative, one widely used indicator of changes is segmentation, so I shall also give a detailed review of the segmentation-based account.

1.3.2.2.1 Memory storage model

Ornstein (1969) had formulated the storage size model; he related the remembered duration of an interval to its corresponding storage size in the memory. Specifically, with an increase in the number of events occurring in an interval or an increase in their complexity, the experienced durations are lengthened (Ornstein, 1969: 41). He also pointed out that storage size is not merely determined by the amount of information stored in the memory, but is also influenced by the way information is stored. In other words, the length of experienced duration also depends on how information is chunked.

Ornstein (1969) carried out a series of studies to support his theory and found that an increase in the number of stimuli, such as consecutive tones, lengthened the remembered duration. He also observed that an increase in the complexity of abstract line drawings (for example, drawings with more angles as opposed to drawings with fewer angles) or tones (for example, each sentence repeated twenty times and then shift to the next sentence as opposed to all sentences presented randomly twenty times) affect retrospective duration judgment. He explained that a greater complexity and increased number of stimuli require larger storage size, which gives rise to longer remembered duration.

Additional evidence in support of this model was provided by Berg (1979) and Mulligan and Schiffman (1979). In Berg's (1979) study, participants were exposed to animations of randomly moving shapes and then undertook a time estimation task. The instructions for the experiment were manipulated so that one group of participants learned the animations as meaningful activities such as people doing something and another group learned the same animations as randomly moving shapes. It was found that the participants who received meaningful descriptions judged the length of the animations as shorter than those who did not. Mulligan and Schiffman (1979) similarly exposed participants to ambiguous line drawings with or without linguistic descriptions. They found that the participants who were given linguistic descriptions remembered the duration of the presentation as shorter than those who were not. These two studies supported storage size theories; specifically, with the help of cues, the participants were able to encode the stimulus more easily and more efficiently which decreased the occupied storage size in the memory and led to a shorter remembered duration.

However, storage size theory is not flawless and there is evidence disproving it. For example, Predebon (1984) found that receiving thematic guidance before being exposed to target prose passages could improve the accuracy of participants' memory but did not influence their time judgment. These results are inconsistent with the storage size theory. Moreover, in a meta-analytic review of twenty experiments, Block and Zakay (1997) found that the number of stimuli may not necessarily relate to time judgments, as the evidence was inconsistent. It should also be noted that the storage size model does not suggest a method to test storage size; it is just a metaphor. Even though researchers manipulate storage size by using stimuli in different quantities or with different complexity, it still cannot equate to storage size because theoretically storage size should be memory-based rather than stimulus-based. Hence, a more measurable model is needed.

Contextual-change model

The change-based model suggests that there are interrelations between contextual change and subjective duration. The model has a long history; James (1890) believed that the perception of time depends on an awareness of changes. Similarly, Fraisse (1963) stated that psychological changes are the basis of experienced duration. Following these ideas, the contextual change model was first formulated by Block and Reed (1978). After many experiments, they concluded that the remembered duration is determined by changes in contextual processing. The more contextual changes there are in an interval, the longer the duration judgment will be for that interval (Block & Zakay, 1996). In line with this view, Block (1982) explored the effects of environment familiarity on duration estimation and found that participants' familiarity with the experiment environment (a laboratory with an experimenter and various objects) decreased remembered durations. He explained this as that when people are exposed to an unfamiliar environment, they encounter more contextual changes which lengthens the remembered durations.

Event segmentation and time

Event segmentation is a measure of the number of changes in an interval. Various studies have shown the importance of event structure or segmentation in representations of time (Avni-Babad & Ritov, 2003; Bailey & Areni, 2006; Block, 1978; Boltz, 1995, 2005; Faber & Gennari, 2015; Jeunehomme *et al.*, 2017; Poynter, 1983, 1989; Zakay *et al.*, 1994). Event boundaries serve as the anchors in organizing episodic memory. It has been shown that the more segments there are in an interval, the longer the judged duration will be (Ornstein, 1969; Bailey & Areni, 2006; Faber & Gennari, 2015). For example, Ornstein (1969) manipulated the number of segments in a dancing movie which comprised 26 movements dividing it into two, six or eleven segments. The results showed that the retrospective duration judgments for the movie correlated with the number of segments; the more segments there were in the movie, the longer the duration judgments. Further evidence for the role of event segmentation was provided by Jeunehomme *et al*. (2018) who tested the number of identified segments for past naturalistic events from their memory (for example, walking to the library) and found that the more recalled segments there were per time unit, the less temporal compression there was of real-life events in duration reproduction.

Faber and Gennari (2015) exposed participants to animations consisting of sub-components of different levels of similarities (similar segments or dissimilar segments) and found that the greater the diversity in the sub-components, the longer the remembered duration was. They suggested that the diversity of sub-events in an interval plays an essential role in remembering and estimating durations of the experienced events.

In summary, in the section 1.3.2.2, we reviewed the cognitive models, which explain psychological time on the basis of the cognitive process. Specifically, the storage size models explain duration judgments by the amount of information remembered whereas contextual-change models and segmentation-based accounts explain remembered durations by the number of changes experienced during an interval. Although the models differ in their explanations of duration judgments, they do not have to be mutually exclusive. As suggested by the EST, encoded event segmentation can modulate the amount of recalled event information (Hanson & Hirst, 1989; Sargent *et al.*, 2013; Swallow *et al.*, 2009; Zacks *et al.*, 2007). The more segments perceived, the greater the recollected information (Hanson & Hirst, 1989; Newtson & Engquist, 1976). Therefore, both the segmentation-based account and the memory-based account can explain duration representation.

Critically, it should be noted that all the models concentrate on the process of encoding and explored event representation by manipulating stimuli. However, the process of estimating or reproducing remembered duration is actually a process of memory recall or memory reconstruction. Hence, what matters in this process is the recall cues and recalled information. Therefore, inspired by memory-storage model, in Chapter 3 and Chapter 4, we would like to propose a recall-based account to explain remembered duration in episodic memory. In addition, we will also suggest an objective measure to test the relationship between recalled information and remembered duration in Chapter 3.

1.3.3 World knowledge, language and psychological time

Our perception, prediction and memory of events are typically guided by our world knowledge or semantic memory. It has been suggested that prior experiences or schematic knowledge also influence duration judgments (Moscatelli & Lacquaniti, 2011; Sgouramani & Vatakis, 2014; Mioni *et al.*, 2015). In this section, I shall review the evidence for how duration estimation is affected by world knowledge.

Moscatelli and Lacquaniti (2011) found that people's experienced duration varied as a function of their knowledge of gravity. They required participants to estimate the duration of an object's movements in four different directions: forwards, downwards, leftwards and rightwards. They found that duration judgments were more accurate for the downward movements, which suggests that knowledge of gravity is automatically activated and facilitates duration judgments. Sgouramani and Vatakis (2014) explored how proficient dancers differ from non-dancers in judging the duration of dancing. They found that dancers had less variability in their time judgment performance compared with non-dancers' performance. Similarly, participants' retrospective duration judgments of naturalistic events were influenced by the variability of event durations. In other words, duration estimations were more accurate for events with a relatively fixed event duration (such as withdrawing money from a cash machine), whereas duration judgments were less accurate when the events did not indicate a fixed duration (for example, a one-stop subway ride). It can therefore be concluded that prior experience and world knowledge can facilitate duration judgment.

In addition, it has also been found that the symbolic meaning of pictures also biases prospective duration judgments. Mioni *et al.* (2015) exposed participants to images indicating either fast speed (for example, a motorbike) or slow speed (for example, a push-bike) and asked them to compare the target intervals with two standard intervals (a standard short interval and a standard long interval); they had to identify whether the target intervals were closer to the standard long interval or the standard short interval. The findings showed that people perceived the faster images as shorter relative to the slower images.

In addition, language also shapes memory of event duration (Loftus, 1975; Burt & Popple, 1996). It was found that post-event information (for example, leading questions) changed participants' memory of the speed of a moving car in a video (Loftus, 1975). The findings showed that systematic errors in memory were induced by people's prior knowledge. Similarly, Burt and Popple (1996) carried out an experiment on eyewitnesses' recall of a staged event (an actor bursting into a lecture theatre). Two weeks later, participants were asked to judge how long the event had lasted. Critically, the researchers manipulated leading questions into a fast version and a slow version. In the fast version of the question, the verb describing the actor's movements suggested fast speed action (such as running) and in the slow version, the verb suggested slow speed action (such as walking). It was found that the event was estimated as shorter if the participants were asked a leading question indicating faster action than when they were given a leading question indicating slow action. This result suggests that being given post-event information about the speed altered people's original memory of the event. Specifically, the participants reconstructed their memory of the event in terms of their schematic knowledge and responded in the way which was appropriate to the question asked of them.

Taken together, remembered duration and perceived duration differ in their underlying mechanisms because they rely on different processing systems (attention as opposed to memory). In addition, duration representation relies on memory, segmentation or changes. Moreover, world knowledge modulates duration estimation and makes a difference in subjective time.

1.4 Putting together time in language and memory

We tend to keep track of event duration in language comprehension and real-world experiences. I have reviewed the literature and theories relating to event representation and time representation. In this thesis, I investigate how we represent event duration in memory and language understanding, and how linguistic concepts or schemas might modulate duration memory. I addressed three questions. First, I examined how we represent event duration in reading comprehension (*see* Chapter 2) by manipulating the length of the events which people read about. Second, I investigate how people encode and retrieve event duration from their episodic memory and how these memory processes can be modulated by schemas conveyed by language (see Chapter 3 and Chapter 4). Third, because event encoding is ultimately determined by segmentation, I asked how prior knowledge of schemas and the specific event structures which they imply modulate segmentation (Chapter 5).

In Chapter 2, I shall follow up previous studies on mental models in language comprehension. Rather than examining time-shifts in narratives, I shall ask how people process and understand temporal adverbials conveying event duration. I used descriptions of the same event which were qualified by longer or shorter temporal adverbials indicating longer or shorter durations (for example, playing basketball for ten minutes as opposed to playing basketball for an hour). Because mentally representing or simulating longer events requires the activation of more diverse or complex event knowledge, as discussed above, it is expected that reading times will increase or decrease accordingly. The chapter thus attempts to provide further insights into the mental models that are constructed during reading. We proposed the world-knowledge plays an important role in building the relevant duration representations in reading, and contrast this approach with embodied approaches.

In Chapter 3, I shall consider the memory processes underlying event memory and duration memory in relation to language. I shall examine how movie stimuli of different durations are encoded and retrieved and whether the remembered duration would be modulated by concepts or schemas associated with the stimuli. Firstly, following storage-size model, I aim to propose a model based on recall process and test the relationship between the amount of recalled information and the length of remembered duration. More specifically, I will measure duration memory by asking participants to mentally replay the video stimuli which could be conceived in different ways using verbal descriptions. In addition, as reviewed earlier in Section 1.2.2, I am also interested in whether and when does language or concept shape remembered duration, at encoding or retrieval stage. On the basis of prior research, I anticipated that duration reproductions would be explained by the amount of information recalled about each video. In addition, I expect that conceptualizations of the videos through language would also modulate duration representations.

In Chapter 4 I shall follow up the findings presented in Chapter 3 by further investigating the situations in which linguistic concepts/schemas can influence memory of event duration. Since the results set out in Chapter 3 suggest that linguistic concepts or schemas do not modulate how stimuli were encoded, I ask whether memory consolidation is required for such modulations to occur. It was expected from prior schema-related memory research that after a period of time, recollections are more likely to be influenced by prior knowledge (schemas). This chapter that further qualifies the interplay between language and memory.

Finally, in Chapter 5, I shall examine whether linguistic concepts varying in the event structure that they imply affect segmentation, the main mechanism for event encoding. As discussed in previous sections, it is known that schematic knowledge contributes to event segmentation. The goal/sub-goal relationships in an event cue event segmentation. However, not all events indicate an explicit aim and the goal/sub-goal relationship might be absent from an event schema. For example, unlike the event ‘cooking potatoes’, the event ‘planes flying in the sky’ cannot be organized into goal/sub-goal relationships. So how do we parse events which do not indicate an explicit aim? Although language schemas were not found to modulate duration memory at encoding, they might nevertheless modulate how events are segmented. In Chapter 5, I shall distinguish events by their implied goals and examine whether explicitly knowing the event aim will affect event segmentation. This chapter thus further addresses the issue of the relation between language and memory, focusing specifically on the segmentation taking place at encoding.

Chapter 2

Representing event duration in stories

2.1 Overview

We all live in time and we talk about time using language. Temporal information is ubiquitous in language (for example, temporal adverbials, tenses and aspect) and readers tend to monitor temporal information during reading comprehension (Rinck & Bower, 2000; Vigliocco *et al.*, 2004; Kelter *et al.*, 2004; Gennari, 2004; Coll-Florit & Gennari, 2011; Zwaan, 2008; Carreiras *et al.*, 1997; Zwaan, 1996). One of the most important types of temporal information is event duration, but it has rarely been investigated. In the present study, I focused on the representation of event duration in language comprehension and explored the underlying mechanisms of duration representations.

2.1.1 Processing events in language

In order to understand the representations of event duration in reading, I shall begin by introducing mental representations of events in language. It has been widely accepted that language comprehension is grounded in background knowledge and prior experience (Van Dijk & Kintsch, 1983; Rinck & Bower, 2000; Rinck & Weber, 2003; Zwaan & Radvansky, 1998; Zwaan, 2008; 1996; Zwaan *et al.*, 2000). So when processing events in language, instead of merely interpreting the text and words with which we are presented, we make sense of what we are reading or listening to by making use of our world knowledge and constructing a mental model to represent the depicted event. For example, O'Brien *et al*. (1998)

found that reading sentences describing events contrary to our world knowledge led to longer processing time. To be specific, if the participants read the sentence ‘Mary has been a strict vegetarian for ten years’ and then read another sentence inconsistent with their predictions (such as ‘Mary ordered a beef burger and fries”), the participants’ reading speed slowed down significantly. This was explained as that the participants did not merely encode the text of stories, but interpreted the stories with their own knowledge (for example, a vegetarian does not eat meat) and formed expectations (such as ‘Mary would not order meat’) about upcoming information. The internal representation of an event is therefore grounded in our world knowledge and prior experience.

Moreover, under the situation model theory, when they are processing language, people automatically construct a dynamic and multi-dimensional mental model (Zwaan & Radvansky, 1998). When we initially encounter a situation, we construct a current situation model which is coded on five indexes: time, space, causation, intentionality and protagonist(s) (Rinck & Weber, 2003; Zwaan, 2008; 1996; Zwaan *et al.*, 2000). Then, when we read about subsequent events, we might encounter changes in any of these five aspects. When this happens, the current mental model has to be updated and the novel information is integrated into it. When we have finished reading the whole text, we store the representations in our long-term memory as a complete event model. Many studies have provided evidence in support of this situation model. They have found that protagonist shifts, spatial shifts, temporal shifts or intention shifts led to a larger processing cost in reading and memory recall (Zwaan *et al.* (1995); Zwaan (1996); Zwaan *et al.* (2000); Rinck & Weber, 2003).

Taken together, there is a consensus that the mental representation of narrative events is grounded in background knowledge and prior experience. In addition, instead of encoding or retrieving the texts and words of narrative events, people construct a mental model to represent the depicted events.

2.1.2 Discourse event duration in language

Temporal information is crucially important for language comprehension. To have a full understanding of a narrative event, readers must not only be aware of when the event occurred but also its duration. Of the five indexes, space, time, motivation, causality and protagonist(s) (Rinck & Weber, 2003; Zwaan, 2008; 1996; Zwaan *et al.*, 2000), time has been shown to play a key role in the internal representation of events, which can also be observed in on-line comprehension and retrieval (Rinck & Bower, 2000; Vigliocco *et al.*, 2004; Kelter *et al.*, 2004; Gennari, 2004; Coll-Florit & Gennari, 2011; Zwaan, 2008; 1996; Carreiras *et al.*, 1997).

There is ample evidence showing that a temporal shift cues updating of the mental model and influences the discourse on subsequent events and access to a previously presented event (Zwaan, 1996; Gennari, 2004; Kelter *et al.*, 2004; Rinck & Bower, 2000). For example, Zwaan (1996) found that relative to a shorter time shift (such as “a moment later”), a longer time shift in a story (“an hour later” or “a day later”) resulted in longer reading times related to subsequent events. These results support the situation model theory that temporal information is involved in mental models. Also, increased discontinuity in the temporal dimension (for example, an hour or a day) leads to greater difficulty in integrating new information into the current model. A new mental model therefore has to be constructed, which results in an increased processing cost for discourse on subsequent events. However, if the temporal discontinuity is relatively small, we can easily incorporate new information into the current mental model and there would be no need to posit a new situation model. Similarly, a greater temporal distance between two consecutive events would make the previously presented event less accessible (Kelter *et al.*, 2004; Rinck & Bower, 2000). Kelter *et al.* (2004) required participants to read stories consisting of two consecutive events. They manipulated the duration of the second event into a long version and a short version ('bake some cookies' as opposed to 'put some cookies') but the first event was identical in both versions. It was found that the reaction time to probes relating to the first event was less accessible (a slower reaction time) if the second event referred to a longer event. They suggested that during memory recollection, people mentally simulate the temporal structure of narrative events. The longer duration of the second event therefore increased the temporal distance from the previous event and decreased access to previous information.

Despite the extensive studies investigating the influence of temporal relationships between events, very few studies have focused on the representation of event duration in a singular event. Coll-Florit and Gennari (2011) found that people needed more time to process durative events (for example, 'owe five euros') than non-durative events ('lose five euros'). In addition, eye movement studies have also provided evidence in support of this view; it was found that relative to listening to sentences indicating longer event durations, listening to events indicating shorter durations (for example, 'The student will run/stagger along the rail to the picnic basket'; 'Stuart hunted/shot a deer') would lead to faster and more eye movements towards the goal/ target (the picnic basket; the deer) (Lindsay *et al*., 2013; Joergensen & Gennari, 2013). In general, the studies investigating duration representation agreed that longer events take longer to process.

However, they explained the duration effects differently and the underlying mechanisms of duration representation in language remain unclear. On the one hand, some studies have explained duration representations in terms of embodied theory. They suggested that people comprehend events by mentally simulating the described events. So longer events take longer to process because people need to simulate the slower motions (Fecica & O'Neill, 2010; Lindsay *et al*., 2013). However, mental simulation fails to explain how people simulate events in a temporally compressed manner. For example, to comprehend the sentence 'playing basketball for one hour' does not take one hour to understand the described event. In contrast to the embodied account, Coll-Florit and Gennari (2011) explained duration representation in terms of the associated background knowledge. They suggested that people took longer to process durative states ('owe five euros') than non-durative events ('lose five euros') because durative states are associated with more diverse knowledge, which leads to greater processing cost. In other words, with more diverse background knowledge activated by durative events, the association strength between background knowledge and narrative events is weaker because the attention allocated to each individual piece of background information is reduced (Hoffman *et al*., 2013).

It should be noted that previous studies which have investigated duration representations of singular events manipulated event duration implicitly by using different verbs (‘*owe* five euros’ as opposed to ‘*lose* five euros’). The observed duration effects were therefore possibly driven by the activation of different event schemas associated with verb meanings rather than by event duration. A better control would therefore be to manipulate temporal adverbials for the same events (for example, playing basketball for one hour as opposed to playing basketball for three hours). In this way, the same event schema will be activated by the verb and the differences between long events and short events in terms of processing cost will be due to the manipulation of the temporal adverbials.

2.1.3 The present studies

In the current study, in order to investigate duration representation in language, I tested how duration information conveyed by temporal adverbials affects event representation. Additionally, previous studies of duration representation only focused on events which had relatively short duration, spanning from seconds (such as switching on a light) to tens of minutes (such as driving to a store), but seldom considered events taking place on longer timescales (weeks, months or years). Many of the events which happen in our lives are, however, of very long duration, such as building a new city. It will be valuable to look at representations of events on a large timescale. Therefore, in the current study, we explored the duration representation for large timescale events and small timescale events to examine whether duration representation differs as the function of timescale.

I also wanted to explore the underlying mechanisms of duration representation and examine whether information activation can explain it. In other words, how the variability or diversity of activated background knowledge predicts duration effects in processing cost. Following the previous findings, it was expected that events of longer duration would be associated with more diverse background knowledge, which would lead to higher processing costs in processing events.

In addition, previous studies have primarily focused on representing the duration (for example, having dinner, cleaning the house) of events with relatively little attention given to duration representation for states (for example, being in love, being tall, being happy). However, linguists and psycholinguists have suggested that differences between states and events in their causal structures lead to different processing costs in language comprehension (Rappaport *et al.*, 1998; Smith, 1991; Taylor, 1977; Vendler, 1967; Verkuyl, 1993). For example, it was found that reading sentences denoting events took longer than reading sentences denoting states (Gennari & Poeppel, 2003). The reason given for this was that events entail a complicated causal semantic structure which led to the observed processing costs. In the current study, I was interested in how people represent durations for states and wanted to test whether the duration effects in processing cost would also be observed for states.

I presumed that the different causal structures associated with states and events would possibly lead to different duration effects in terms of processing cost. Events involve dynamic changes from an initial state to the final state, so more changes would be expected to occur in events with longer durations. This would activate more diverse and a larger amount of information when we process longer events, which would increase the processing cost. Nevertheless, states emphasize a stable status without suggesting changes occurring in the interval (Rappaport *et al*., 1991; Taylor, 1977; Vendler, 1967; Verkuyl, 1993). So with both longer and shorter durations, the activated information for states might be unchanged and processing costs might remain the same regardless of a state's duration. It is still unclear whether the duration of state stories will indeed change processing cost. I therefore carried out an experiment to explore duration representations for state stories.

Following the previous work, in the current study, I would like to carry out experiment to address three main questions: 1) how people process durations of large-timescale events and small-timescale events in language (Experiments 1 and 2); 2) the underlying mechanisms of duration representation (Experiment 3); and 3) duration representations for states (Experiment 4).

To address these questions, in Experiment 1, I reported a previous experiment of small-timescale story and constructed event stories on a large timescale (events with duration more than one day: days; weeks; months) and tested participants' reaction time to probes associated with longer events or shorter events in large timescales ('It took three months to build a wall' as opposed to 'It took two weeks to build a wall'). In Experiment 2, I combined stories on a large timescale (more than one day; days; weeks; months) with stories on a small timescale (within one day; minutes; hours) and examined how the processing cost to probes changed with timescale. In Experiment 3, in order to reveal the underlying mechanisms of duration representation and investigate the role of activated background knowledge in explaining duration representation, I collected participants' free recall of the stories and analysed the variability of the content of their answers. Finally, in Experiment 4, in order to explore duration representation in state stories, I constructed state stories (for example, knowing somebody for one year; two years) with the same story structure as the event stories and tested the reaction times to probes associated with the target state stories.

Following the previous findings, it was expected that duration effects would be seen in large-timescale events. In addition, large-timescale events should take longer to process relative to small-timescale events. It was also expected that activated background knowledge can explain duration representations. In other words, relative to a short event, longer events should activate background knowledge of greater diversity or higher variability. Also, as discussed earlier, due to the difference in causal semantic structures associated with states and events, the duration effect might possibly be absent in state stories. However, it is still unclear whether the duration of state stories will indeed change processing cost. This will be left as an open question for now.

2.2 Experiment 1: Representing duration in large-timescale events and small -timescale events

In the first experiment, I sought to explore duration representations of large-timescale events and small-timescale events. Previously, we have examined duration representation for stories with durations of hours minutes (all within one day) (as reported below). In the current experiment, I constructed stories with durations of days, weeks, months and years (all beyond one day) and manipulated the duration of temporal adverbials and probe positions. It was predicted that on a large timescale, longer events would lead to higher processing costs and people would take longer to respond to probes associated with longer events.

2.2.1 Previous work of small-timescale events

In a previous study, we explored representations of event duration by manipulating temporal adverbials (see below). We constructed stories containing two events and the second event was manipulated into a longer version and a short version by using temporal adverbials indicating longer or short durations (for example, an hour; three hours). The first event therefore remained unchanged in the long and the short versions, whereas the second event (the critical event) was either of a longer or a shorter duration in the two versions of stories. In this experiment, participants read the stories sentence by sentence (as shown below in Story Example) and then immediately after reading each story, they were required to judge whether a word (a probe) was presented in the story. The probes were designed to be associated with either the first event (early probe) or the second event (second probe). The late probes were therefore associated with events varying in duration whereas the first probes were associated with an event without any duration manipulation. After judging the probes, they were given comprehension questions referring to the content of stories, so they had to focus not just on the words but also on the meaning and the content of the text.

Reaction times and accuracy to probes were recorded. It was expected that duration effects in reaction time would only be observed for the late probes because late probes were associated with either longer or shorter events in the two versions of stories. However, the results were inconsistent with the prediction because there were no significant duration effects in reaction time for either early probes or late probes.

Story Example:

- Small-timescale story example (days, weeks, months)

John wanted to impress his wife.

He had a few hours/an hour free before his favourite TV show.

He used all that time to sand her desk.

Then, he watched the show.

Early Probe: ‘impress’ Late probe: ‘desk’

Comprehension question: Did John spend an hour sanding a desk?

The absence of duration effects was possibly due to the fact that the difference in event duration between the short condition and the long condition in the experiment was not salient enough. Duration effects might have been seen if larger duration differences were introduced. Following this previous study, in the present study I constructed stories on large timescales (days, weeks, months) and explored the duration representations for large-timescale events and examined how duration representations varied as a function of event timescale. As explained above, it was expected that the duration effects in processing cost would be seen in large-timescale events. In addition, as events on a larger timescale usually relate to a larger number of sub-events and more diverse background knowledge, it was expected that people would take longer to process large-timescale events than small-timescale events.

2.2.2 Method

Participants

I recruited 76 native English speakers students at the University of York and each received a small cash payment in recognition of their participation. Of these participants, 52 were recruited during my master's study period, which had been reported in my Master thesis. Additional 24 participants were recruited in the first year of my PhD study, therefore, the data analysis in Experiment 1 was based on the larger set of data from that previously used.

Materials

I used 28 experimental story pairs previously constructed by Silvia Gennari. The story pairs were of large-timescale events (more than one day) and had the same story structure. Each story pair had a short version and a long version (two weeks VS. one month).

In greater detail, each story was composed of four sentences and constructed following the same structure, as shown below. The first sentence described the general background knowledge of the story. And then, in the second sentence, time adverbials were introduced referring to longer or shorter durations (one month vs. two weeks). The third sentence further described what the protagonists of the story did during that period of time. The final sentences were consequences or additional activities. Importantly, it should be noted that the first sentence appeared before the temporal adverbials, and the described information in the first sentence did not have a specific duration. However, the third sentence depicted an event with a specific duration, and this was considered to be the critical sentence. The two versions of the stories were therefore identical except for the duration of the critical events. In addition, to rule out any possible compounding effect due to tense, all of the stories were presented in the past tense. The plausibility of the stories was tested in advance on a 7-point scale using an online norming questionnaire. No significant differences in plausibility rating were found between the short and the long condition (the mean number of plausibility rating for long condition and short condition was 3.95 vs. 4.12 respectively).

Two probes were created for each story. The early probe was a word appearing in the first sentence which was associated with the background scenario (for example, ‘government’). The late probe was a word appearing in the third sentence (for example, ‘build’) and was associated with the critical event. The late probe therefore related to an event varying in duration in the two versions of the stories, whereas the early probe was not influenced by duration manipulation. A comprehension question referring to the content of story was also constructed.

In addition, 42 filler stories were included which were similar to the experimental stories. The answers to the probe recognition task and comprehension questions were counterbalanced, with half Yes answers and half No answers. All the stories (including filler and experimental stories) were presented to participants in random order.

Story Example:

- Large-timescale story example (days, weeks, months):

Several prisoners had escaped and the government was furious.

The prison director was given ***one month/two weeks*** *to improve security.*

It was just enough time to build a wall.

The director was pleased with the result.

Early probe: ‘government’ Late probe: ‘build’

Comprehension question: Was a wall built at the prison?

Overall, for each experimental story there were four versions, ‘early-short’, ‘early-long’, ‘late-short’ and ‘late-long’ depending on the probe positions and the duration of events respectively.

- *Design and procedure*

Four versions of 28 experimental story pairs were allocated to four lists in a Latin square design, so there was one version of each story in each list and the conditions of the stories in each list were equal in number. Each participant was therefore presented with 28 different experimental stories and 42 filler stories and read the same numbers of stories in the ‘early-short’, ‘early-long’, ‘late-short’ and ‘late-long’ conditions. In addition, all the stories were presented randomly, and participants were randomly allocated to one of the lists.

In the experiment, texts were displayed on a computer screen sentence by sentence using the moving-window paradigm. By pressing a button on the button box, participants would see the next sentence of the story. Immediately after reading the whole story, there were two questions relating to the previous story. The first question was a probe-recognition task asking them to judge whether they had seen this word in the previous story. Then they were given a comprehension question about the story. They were required to indicate their responses by pressing the buttons labelled Yes or No. They were given two practice trials before the experiment started and in order to avoid the participants giving deliberate attention to temporal information, they were unaware of the aim of the study during the test. After completing the whole experiment, the participants were debriefed about the aim of the study.

- *Data treatment*

Only correct responses to the experimental stories were used for the data analysis. Outliers of reaction time were determined separately for each condition. To determine outliers, a Z-score was calculated for each trial and any Z-score which deviated more than 2.5SDs from the mean of each condition was excluded. This process excluded 5% of the trials.

2.2.3 Results

- *Accuracy of probe recognition*

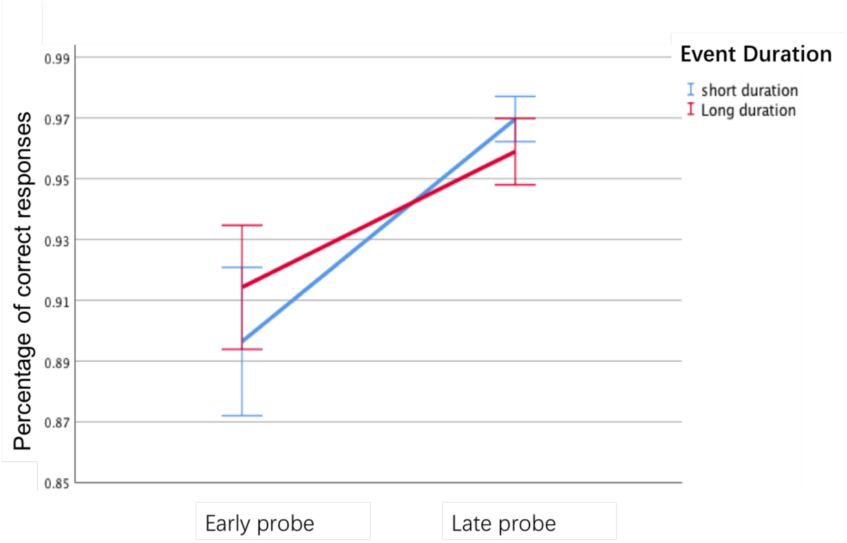


Figure 2.1 Experiment 1, Accuracy: Response accuracy to early and late probes in large-timescale stories in long and short conditions. Error bars indicate standard error.

A 2X2 repeated ANOVA with the proportion of correct responses to probe recognition as the dependent variable was performed. There were two independent variables: probe position (early, late) and duration (long, short). The main effects of probe position were observed ($F_1(1,75) = 29.15$, $p<.001$; $F_2(1, 27) = 8.10$, $p = 0.008$) and there were no other main effects or interactions (See Figure 2.1). The proportion of correct responses to the probe recognition question for early probe and late probe were 0.905 and 0.964 respectively. Responses to the early probes were less accurate than those to the late probes as the words being probed were presented at the beginning of the stories and were therefore less accessible.

- *RTs to probes*

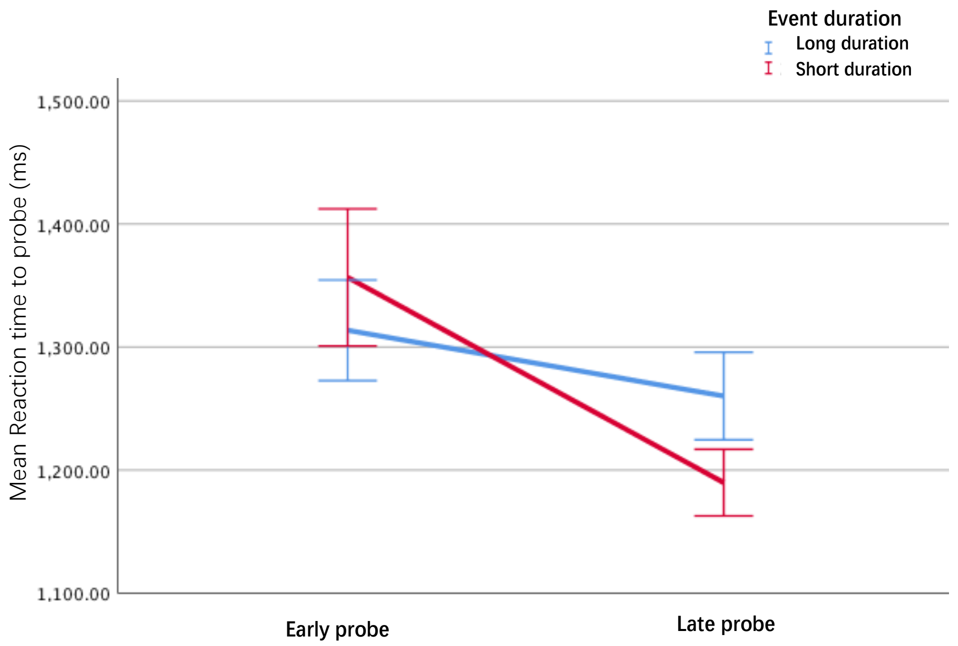


Figure 2.2 Experiment 1: Response time to early and late probes in large-timescale stories in long and short conditions. Error bars indicate standard error.

Table 2.1: Experiment 1***:*** Mean reaction time to probe and standard deviations

Probe position	Condition	
	Long duration	Short duration
Early probe	1313.58(354.02)	1356.57(482.57)
Late probe	1260.15(308.33)	26.85(234.11)

Standard deviations are given in parentheses.

A repeated measure ANOVA was carried out with transformed reaction time as the dependent variable and duration (long, short) and probe position (early, late) as independent variables. The analysis of reaction time to probes showed a main effect of probe position ($F_1(1,75) = 15.62$, $p<.001$; $F_2(1,27) = 6.89$, $p = .014$). Because the early probe was presented earlier in the text, it was reasonable that it would take longer to recover the information from memory.

An interaction was also found between duration and probe position in reaction time (F1(1,75) = 4.45, p = 0.038; F2(1, 27) = 6.24, p = 0.019), as shown in Figure.2.2. To examine this interaction further, a paired T-test was performed and no significant difference was found between the long and the short condition in reaction times to the early probe (p>.05), whereas the reaction times to the late probe showed a significant difference between the long and short conditions (t1(75) = 2.74, p = 0.008; t2(27) = 2.14,p = 0.041) that reaction times to late probe were significantly longer in long condition than in short condition. The means and standard deviations for reaction times to the early probe and the late probe were given in Table 2.2.

2.2.4 Discussion

In Experiment 1, consistent with the predictions, duration effects were observed in reaction time for the late probe in the large-timescale stories. The participants took longer to recognize a probe associated with a longer event whereas early probes did not vary in reaction time because they referred to events with no differences in duration or in its temporal distance to later critical events. Indeed, unlike prior studies showing effects of temporal distance between events (e.g., Kelter et al. 2004), the first sentence in our stories could overlap later events, for example, the state of the government being furious in our example above might overlap with the wall building, rather than entirely precede it in time. Thus, responses to the early probe did not vary with the duration manipulation. The main effects of probe position (the early probe took longer to retrieve) can be explained by recency. Specifically, the last-mentioned information in stories can be retrieved faster than information given earlier (Glenberg & Swanson, 1986).

The results were consistent with the predictions and with the findings of previous studies of duration representations in that longer events took longer to process. However, duration effects were absent in small-timescale events referring to minutes or hours long events, as was observed in our previous work (see sec.2.2.3). The inconsistent duration effects should not be due to statistical power as there were the same numbers of stories and participants in the two experiments. One possible reason for the inconsistent duration effects was that the duration differences in the large-timescale stories between long and short conditions were also larger, compared with the duration difference in small-timescale stories. Another possibility was that, events of large timescales (days, weeks, months) included more sub-events punctuated by overnight sleep of the main characters, which segments the represented duration and brings to mind many different situation models, which may increase the processing cost. Such internal segmentation was absent from small-scale events (minutes and hours long).

In the second experiment, duration representation between large-timescale events and small-timescale events was compared. Large-timescale stories and small-timescale stories were presented to the same group of new participants and their responses to the two types of story were examined. As explained in the introduction, it was expected that the timescale effects in reaction time would show that large-timescale events would take longer to process than small-timescale events.

2.3 Experiment 2: Timescale effects in stories

The purpose of the second experiment was to investigate how people construct mental representations of events of different timescales (within one day as opposed to longer than one day). To achieve this, the small- and large-timescale stories were combined in one experiment.

As in the previous experiments, long and short durations within each of the time scales were used. Consistent with previous findings, it was anticipated that the timescale effects would be observed, i.e., large-timescale stories (days, weeks, months) would take longer to process than small-timescale stories (minutes, hours). Besides, it was also expected that the duration effects previously found in the large-scale stories would be replicated.

It should be explained that the previous stories were combined into one experiment but the stories were not changed. The intention was to use the same verb/event with all possible temporal adverbials (one hour, two hours, one week, two weeks). However, it was difficult to find verbs and actions which could be used in all these different durations and maintain the plausibility of the stories. Typically a meal is not cooked in a week or a month, and bridges are not built in minutes. Such combinations would lead to less plausible temporal adverbials across conditions, so this feature was controlled in the stories.

2.3.1 Methods

Participants

The same 73 students from the University of York participated in return for either a course credit or a smell cash gift. All of the participants were native English-speaking students. One participant was excluded because his/her low accuracy in the probe recognition (80%), and comprehension questions (77%) greatly deviated from the mean group performance (around 90%). The experiment was approved in advance by the Ethics Committee of the Department of Psychology at the University of York.

Materials

The stories of both timescales (See Example Story presented in Experiment 1) used in the previous experiment were used again and were combined into one experiment. Bearing in mind the length of the experiment (approximately 50 minutes), sixteen of the 56 experimental stories and sixteen of the 56 fillers from the materials of the previous experiment were discarded. There were therefore 80 stories (twenty large-timescale stories, twenty small-timescale stories and 40 fillers) in the second experiment.

As in the previous work, each experimental story was manipulated into a long version and a short version and the probe position was also manipulated. An early probe appeared at the first sentence and was not related to the temporal adverbial, and a late probe occurred at the third sentence and was related to the temporal adverbial. In addition, the same comprehension question was used in all four conditions of each story; this was again a Yes/No question relating to the content of the story. Importantly, all the probe recognition questions for the experimental stories should receive a Yes response.

In addition, to counterbalance the Yes and No responses to the probe recognition questions throughout the experiment, the 40 fillers were included so that there were equal numbers of Yes and No responses across the entire stimulus set. The answers to the comprehension questions were also counterbalanced so that half of them should receive a Yes response and the other half a No response.

Design and procedure

The items were arranged into four lists, each containing all 40 experimental stories and 40 filler stories. In addition, for each list there were five items in each of the four conditions (short/long small-scale and short/long large-scale). Again, the items were allocated to the different lists using a Latin square design which meant that each participant only saw a given item in one experimental condition but saw all conditions across all items. E-Prime2 software was used to present the stimuli and collect the responses.

The same procedure was used as for Experiment 1. In Experiment 2, the participants were instructed to read the stories sentence by sentence at their own normal reading speed; they were then required to take the probe recognition task and answer the comprehension question.

Data treatment

Only stories which produced correct answers to the probe question were included in the analysis of response times: 93% of the responses were considered to be valid. Additionally, following my previous work in my Master thesis and Experiment 1, response times to the probes which were more than 2.5 standard deviations from each condition mean were excluded from data analysis, and this removed 3.1% of the data. In total, therefore, 90% of the responses were subjected to further analysis.

2.3.2 Results

Accuracy of probe recognition

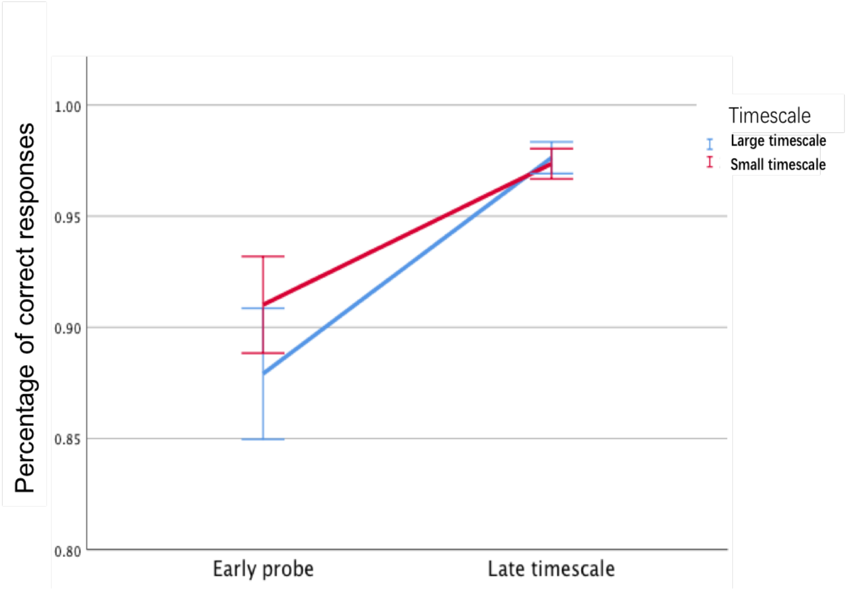


Figure 2.3： Experiment 2，Accuracy: Response accuracy to early and late probes in large-timescale stories in long and short conditions. Error bars indicate standard error.

A 2X2X2 repeated ANOVA was performed with the proportion of correct responses to probe recognition as the dependent variable; there were three independent variables: probe position (early, late), duration (long, short) and timescale (large, small). The main effects of probe position were noted ($F_1(1,71) = 68.81$, $p<.001$; $F_2(1, 19) = 15.07$, $p = 0.001$) and there were no other main effects or interactions (see Figure 2.3). The proportions of correct responses to the probe recognition question for the early and late probe were 89% and 97% respectively. As expected, the early probes produced less accurate responses than the late probes because they appeared at the beginning of the stories and were therefore less accessible.

RTs to probes

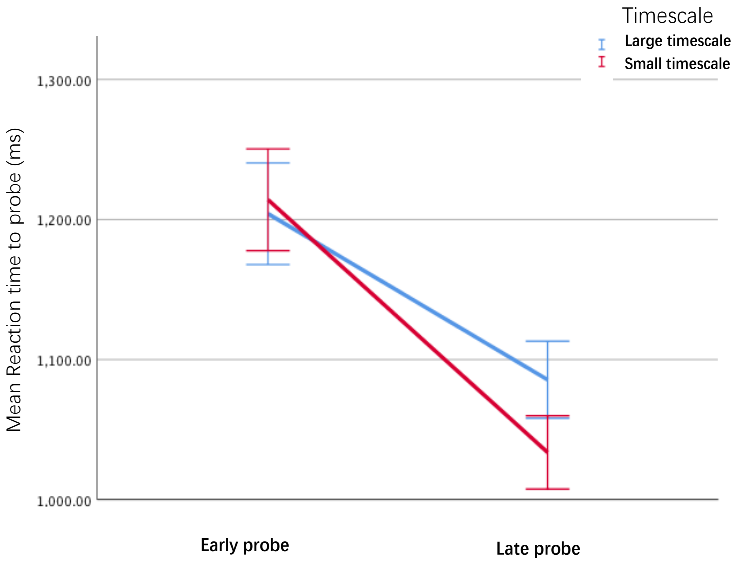


Figure 2.4 Experiment 2: Response times to early and late probes in large-timescale stories and small-timescale stories. Error bars indicate standard errors***.***

Table 2.2 Experiment 1: Mean reaction time to probe and standard deviations

Probe position	Large timescale		Small timescale	
	Long duration	Short duration	Long duration	Short duration
Early probe	1186.73(302.73)	1207.36(370.79)	1217.74(335.88)	1214.55(370.80)
Late probe	1072.96(233.01)	1098.07(256.83)	1048.06(239.06)	1035.85(226.12)

Standard deviations are given in parentheses.

A repeated ANOVA with RTs to probes as the dependent variable and timescale (large, small), probe position (early, late) and story duration (long, short) as independent variables showed the main effects of probe position ($F_1(1,71) = 53.07$, $p < 0.001$; $F_2(1,19) = 25.15$, $p < 0.001$). An interaction was also found between probe position and timescale ($F_1(1,71) = 7.09$, $p = 0.01$; $F_2(1,19) = 4.88$, $p = 0.04$). There were no other main effects or interactions observed. In order to explore further the interaction between probe position and timescale, paired T-tests were performed and it was found that large timescale and small timescale significantly differed at late probes in both by-subject and by-item analyses ($t_1(71)$

= 4.71, $p < 0.001$; $t_2(19) = 3.42, p = 0.003$) and there were no differences between timescales in response times to the early probe ($p > .1$), as shown in Figure 2.4. The means and standard deviations for reaction times to probe are given in Table 2.2.

2.3.3 Discussion

The aim of the second experiment was to test how duration representation varies as a function of timescale. The findings showed that relative to stories of small timescale (minutes, hours), stories of large timescale (days, weeks, months) took longer to process. Interestingly, these effects were only observed in the response times to late probes. The results were consistent with the predictions and the findings of previous experiments indicating that events of longer duration were associated with greater processing cost. One may argue that possibly the timescale effects were due to the use of different stories in small-timescale stories and large-timescale stories. However, the characteristics of the word forms are unlikely to confound the results, because we measure working memory processes in these reading studies, rather than word retrieval anew. In the experiments, the words have just been read in the stories and probe responses target the difficulty in accessing these words. In this situation, probe responses represent the degree of accessibility of these previously seen words. Thus, words properties like frequency would play a relatively minor role compared to the cost of retrieving the content of the story.

However, we did not find duration effects within a time scale (long vs. short) in the present experiment, which was not in line with previous findings for the large-scale stories. One possibility is that the materials used in the experiment did not have enough statistical power. In the present experiment, we reduced the number of stories previously used from 28 to 20, each with a long and a short version. Another possibility is that we lost sensitivity in the present experiment by combining the large- and small-scale stimuli. As discussed above, the differences in duration across conditions is larger across scales than across short or long versions within a scale. Therefore, it is possible that the experimental context made the scale duration differences (hours vs. weeks, months) more salient than that within a scale (one hour vs. two hours).

Another possibility is that the scale effects were driven by the verb semantics and their situation models, rather than the temporal adverbials themselves. Recall that the large timescale stories and small timescale stories used different verbs (build a wall vs. sand a desk). They differed not only in the event durations but also possibly on other aspects. Although the number of words in the story and the number of letters in the probes did not differ across stories, they did differ in the verbs and the situation models activated (e.g., sanding a desk/drawing a tiger vs. erecting a wall/assembling a boat). As suggested above, these events may imply different internal segmental complexity, and therefore, simply re-activating the corresponding event at the probe may lead to different response times. In other words, differences across scales are differences in duration, but these differences may be more like the lexically driven differences discussed previously (e.g., owe money vs. lose money), rather than differences driven by the temporal modifications themselves.

Nevertheless, previous results suggested that there were long vs. short duration effects for large-scale events (one week vs. two weeks). Thus, the fact that these effects were not observed in the present experiment may be due to a combination of factors: reduced number of items (lack of a statistical power), larger lexically driven duration differences present in the experimental set, and the fact that the small-scale items were not likely to differ in long/short durations as previously shown, thereby reducing the likelihood of an overall main effect of duration.

In Experiments 1 and 2, duration representations in events were explored and timescale and duration effects (long, short) were found in large timescale events. The purpose of the next experiment was to explore the underlying mechanisms of duration representation. As was explained in the introduction, duration representation is possibly driven by the diversity or variability of the associated world knowledge, so in the third experiment, the activated background knowledge in processing long events and short events was tested and the diversity or variability of the activated background knowledge was analysed. It was expected that greater variability in background knowledge would be found for long events than for short events.

2.4 Experiment 3: the underlying mechanisms of duration effects in language processing

In Experiments 1 and 2, it was found that longer events took longer to process. Coll-Florit and Gennari (2011) found that longer events were associated with more diverse background knowledge which led to greater processing cost and longer processing time for long events; the results showed the crucial role of background knowledge in explaining the observed duration effects in language. The current study was designed to examine whether variability of background knowledge can explain the observed duration effects elicited by using different temporal adverbials.

In this experiment, diversity or variability of background knowledge was defined as the extent to which the activated knowledge of a story differs in event type. To be more specific, if the activated knowledge of a story is dominated by a single event type (e.g., mental state vs action-based events), we would consider the activated knowledge as less diverse. On the contrary, if the activated knowledge of a story was less dominated by one type of association and evenly composed of different event types, then we would consider the background knowledge as being more diverse. To test the diversity of background knowledge, following Coll-Florit and Gennari's (2011) study, the diversity of the elicited associations was measured by calculating the difference between the proportions of event associations and state associations elicited by the stories. Because states and events differ greatly in their internal causal structure and temporal developments, a larger difference in the proportions would show that most associations are of the same type or less diverse. Alternatively, when there is a smaller event/state difference, the associations are expected to be more balanced across the semantic types, which would mean that they would be more diverse in semantic type. Consistent with the previous findings, it was expected that the longer events would associate more diverse background knowledge. Moreover, we also expect to observe a correlation between processing time and variability of background knowledge.

In this study, first the associations relating to the stories were collected and then the proportions of state associations and event associations were calculated based on the verbs appearing in participants' answers. Only large-timescale stories were used to see whether different temporal adverbial would elicit different background knowledge and explain the observed duration effects in processing cost.

2.4.1 Methods

Participants

A total of 36 native English-speaking students were recruited from the University of York; their participation was rewarded by a small cash gift or a course credit. Two participants were excluded from the data analysis because they provided very limited associations, so data from 34 participants were analysed.

Materials, design and procedure

Two questionnaires were used which included all of the 28 large-timescale stories used in the previous large-timescale experiment. The long and short versions of the stories were allocated to different lists and the numbers of short- and long-version stories in each list were counterbalanced.

The process followed the same structure as the previous experiments. The participants read each story and were then asked to describe what would happen during the duration of the critical event in the story based on their intuition. For example, 'What do you think happened during the time period in which Matt was fixing the truck?' To answer this question, they were instructed to consider the story context and the things which they thought likely to occur, using their intuition. They were also instructed to provide as many answers as they could easily and intuitively think of. They were required to type their answers to the question in a blank box.

Data coding

The coding was based on the verbs appearing in the participants' answers. All the verbs and verb phrases were coded as either 'event' or 'state'; verbs or verb phrases indicating a simple action or a group of actions were coded as events (for example, "have dinner, write, help somebody"), and verbs which indicated a stable status and cognitive activities referring to no physical action were coded as states (for example, "being happy, being generous, thinking about something"). Typically, emotional and mental verbs were considered as states (for example, "like, enjoy, be frustrated, miss home") whereas physical actions or events including active agents, even if the physical actions were not specified, (for example, "help somebody, increase security at a prison, research tourists' sites") were considered as events.

For each story**,** the proportion of events and states mentioned out of the total of verbs used across participants for that story were computed. All of the responses (from the twenty participants in each group) were considered as a whole when computing the proportions for a given story. Across the story set, there were no significant differences in the total number of verbs considered for each short or long condition ($t(33) = 1.4$, $p > 0.05$). In total, there were 1844 verbs in the long condition and 1930 verbs in the short condition. The difference in the proportions of event associations and state associations were calculated for each story.

2.4.2 Results

Consistent with the findings of prior studies which examined lexically driven duration ('lose money' as opposed to 'owe money') (Coll-Florit & Gennari, 2011), more event associations were found than state associations (82% event associations and 18% state associations). This finding was reasonable, as the stories described events composed of sub-events, so the most related information would be the associated sub-events. Moreover, the question 'What might happen in the event' also cued event association answers.

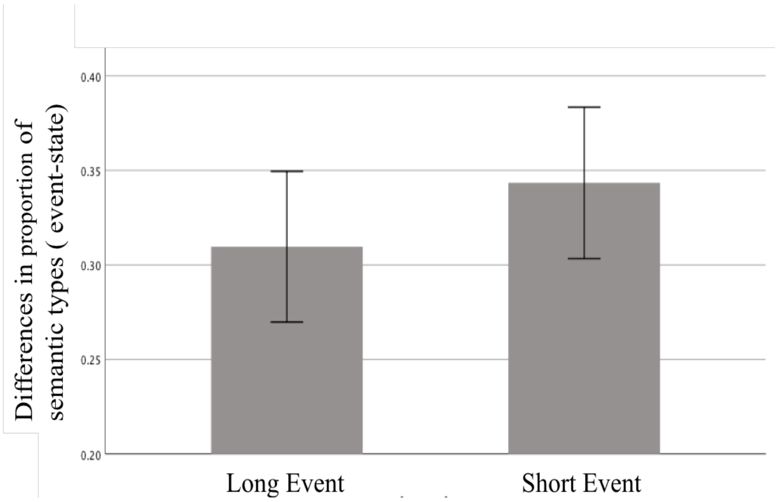


Figure 2.5 Experiment 3: The difference in proportion between event associations and state associations for long events and short events in large timescale. The error bars indicate standard errors.

Importantly, consistent with the findings of Coll-Florit and Gennari (2011), it was found that the differences in the proportion of state associations and event associations varied as a function of event duration, given that the lexical and sentential content of the stories was identical except for the temporal adverbials. Paired T-tests were carried out and the differences between event and state proportions for long and short stories were found to differ significantly both by item and by subject ($t_1(27) = -2.08$, $p = 0.047$; $t_2(33) = -3.01$,$p = 0.004$). This finding indicated that short stories had less varied associations in comparison with long stories (*see* Figure 2.5). Because the content of the stories was the same, even though the difference was relatively small (long story: short story – 0.34:0.30), the significant results indicated the contribution of the temporal adverbials in activating different background knowledge.

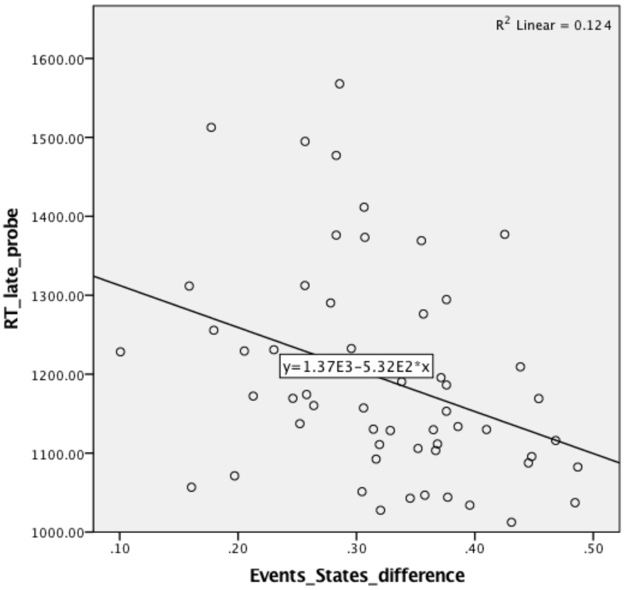


Figure 2.6 Experiment 3: Correlation between reaction time to late probe and event/state proportion.

Table 2.3 Experiment 3: *Step-wise regression model*

Model	Factors	Unstandardized B	Coefficients Std. error	t value	sig
1	Constant	1197.89	17.06	70.20	.00
	Diversity of background knowledge	-73.95	27.38	-2.70	.009
2	constant	1237.587	55.09	22.47	.00
	Diversity of background knowledge	-69.00	28.25	-2.44	.018
	Duration	-25.63	35.11	-0.758	.452

Consistent with the findings of Coll-Florit and Gennari (2011), a negative correlation was found between the event-state differences and processing time. When this difference became smaller, the background knowledge would be more distributed between events and states or more diverse in content. There would be larger processing costs or longer response times to the probes. In contrast, when the event-state difference was larger, associations were less distributed in semantic types and there would faster reaction to the probes. In line with the predictions and with previous findings, a negative relationship was observed between event/state proportion differences and reaction times to the probes (R = 0.35, p = 0.008) (*see* Figure 2.6, Table 2.3).

To ensure that the contribution of event-states indeed accounted for most of the variance in the reaction times to probes, over and above the duration grouping condition, a multiple step-wise regression was conducted with RTs as the dependent variable. After entering the event-state difference into the model, adding the conditions (short, long) – coded with dummy variables – into the model did not increase the proportion of variance observed. In other words, the duration effect in processing cost can be explained by the variability of background knowledge as much as by the coding conditions (short, long).

2.4.3 Discussion

The aim of the third experiment was to explore how the activated background knowledge during language representations varied as a function of event duration and whether the difference in activated background knowledge would predict the duration effects observed in the previous experiments. Consistent with the predictions and with the previous findings, significant differences were observed in the variability of the background knowledge associated with the long and the short stories. Moreover, the difference in background knowledge predicted processing time, showing that the observed duration effects in processing costs were driven by the variability of activated background knowledge.

The results indicated that longer stories activate more diverse knowledge whilst shorter stories activate less diverse knowledge. The findings also showed that the variability in semantic types of background knowledge explained the processing cost, shown as the correlation between diversity of background knowledge and processing times to the probes. This explanation is consistent with models of semantic memory and the retrieval cost of concepts. The more varied associations linked to a concept, the less activation an individual association has, leading to more difficult retrieval due to retrieval effort or competition between different associates (Hoffman, 2013). Following these ideas, for events of longer duration, possibly participants needed to activate more diverse and weaker associated events simultaneously from the situation model just read, leading to greater retrieval effort. However, for events of shorter duration, the associated information was less diverse, so that the activation level of the individual associations was higher and participants could recover the situation model more easily to respond to the probe. Although more research is needed to confirm this, it does for the first time give an explanation of duration effects which result from temporal adverbials rather than from lexical meaning. Importantly, this explanation is consistent with the duration effects resulting from lexical meaning and thus points to common mechanisms underlying representations of duration in language comprehension.

The previous experiments focused on the representation of duration in events and revealed the underlying mechanisms for event duration. However, how do we represent durations for states (for example, being happy) which highlight a stable status rather than changes and causal structures? Can similar duration effects be observed in the processing cost for a state story? The next experiment was carried out to address these questions.

2.5 Experiment 4: representing duration information in a state story

In the previous experiments, the participants' mental representations of events with different durations were explored and it was found that longer events took longer to process. This effect appeared to be due to the fact that more diverse knowledge was associated with long events. If this was indeed the case for events, what should we expect for stories referring to states?

As already explained, states entail stable statuses rather than changes and causal contingencies whereas events entail dynamic changes from an initial state to a final state. In addition, the previous experiment showed that the activated background knowledge for events varied as a function of duration, and more diverse information would be activated when processing longer events, which would increase the processing cost. Nevertheless, states emphasize stable status without suggesting the possible changes which might occur in the period (Rappaport *et al*., 1991; Taylor, 1977; Vendler, 1967; Verkuyl, 1993). So with longer or shorter durations, the activated information for states might remain the same and processing costs might not be influenced by changes in states' duration. However, it is still unclear whether temporal adverbials will indeed change the processing cost for state stories. The fourth experiment was carried out in order to address this question.

2.5.1 Methods

Participants

A total of 74 native English-speaking students were recruited from the University of York; they were rewarded for their participation by a small cash gift or a course credit. Six participants were excluded from the data analysis because of the low accuracy in their responses (less than 80% in the probe recognition task), so data from 68 participants were analysed.

Materials and item pre-test

Twenty-four state stories with large timescale durations (days, weeks, months) were constructed. The story structure was similar to the event stories used in the previous experiments but not identical. Using story structures identical to those used for the event sentences was not possible because it is implausible to say that anyone spent time in a given state. Recall that in the critical event sentences, phrases such as ‘They spent that time erecting a wall’ were used. A state verb in this context would either be nonsensical (for example, ‘They spent that time knowing the truth’) or the context would force an active event interpretation in which the referred state is actually intentionally or actively pursued (for example, ‘She spent that time being happy’). If event-like interpretations are encouraged by the context, it would not be clear whether states are indeed processed differently or whether the results are caused by having more active event-like interpretations.

Each story consisted of four sentences as shown below. The third sentence introduced a state (for example, knowing somebody, being happy) of either long duration or short duration by manipulating the temporal adverbials (the critical sentence). The fourth sentence then further explained the story. It should be noted that the first sentence presented before the temporal adverbial was not linked to the time of the adverbial. Additionally, as in the previous event experiments, two probes in different positions were devised, early and late. The early probe appeared in the first sentence, the late probe appeared in the third sentence and was related to a duration change. One comprehension question about the content of the story was again prepared.

In total, each story was manipulated into two versions (long and short) in terms of the duration of the critical state and each story had two probes (early and late). Again, 24 fillers were included in the experiment to counterbalance the Yes and No responses.

- ***Example of state stories***

Tony's dog would not eat or move.

Beth recommended the neighbourhood vet.

She had known him for <u>a year/ two years</u>.

And he had a good reputation.

Early Probe: 'eat' Late probe: 'known'

Comprehension question: Had Beth known the vet for a year?

The stories were pre-tested for plausibility to ensure that the different temporal adverbials did not result in more or less plausible sentences. To this end, an online questionnaire was used to ask 30 participants to indicate how likely it was for someone to know someone else for a year/two years, in the case of the example above. The participants indicated their response on a scale of 1 to 7 (‘very unlikely’ to ‘very likely’). No differences were found between the long and short versions of the stories ($p > .05$).

Design and procedure

The experimental design was the same as in the previous experiments. Items were arranged in four lists, each containing all 48 stories, and in each list, there were six items in each condition. The different conditions of each item were allocated to different lists using a Latin square design, so each participant only saw one version of a story but an instance of all stories. E-Prime2 software was again used for presenting the stimuli and collecting the responses.

The procedure was the same as in Experiment 1; the participants first read a story sentence by sentence and then carried out the probe recognition task immediately after reading the whole story. They were then given the comprehension question. They were required to give Yes or No answers by pressing the appropriate buttons.

Data Treatment

Only responses which gave correct answers to the probe recognition question were included in the analysis of response times; this meant that 93% of the responses were considered to be valid. Additionally, response times to probes of more than three standard deviations from the mean were excluded from the data analysis; this removed a further 5% of the data. In total, therefore, 89% of the responses were subjected to further analysis.

2.5.2 Results

Accuracy to probe recognition

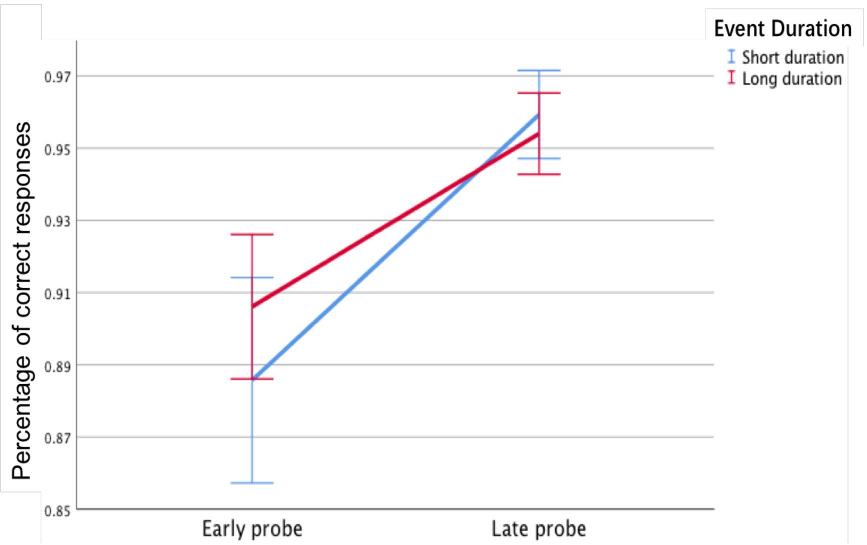


Figure 2.7 Experiment 5, Accuracy: Response accuracy to early and late probes in large-timescale stories in long and short conditions. Error bars indicate standard error.

Repeated ANOVAs were carries out with the proportion of correct responses to the probes as the dependent variable and probe position (early, late) and duration (long, short) as independent variables. Only the main effect of probe position was found ($F_1(1,67) = 12.91$, $p = 0.001$; $F_2(1,23) = 5.194$, $p = 0.034$) and no other main effect or interaction was found (see figure 2.7). As expected, early probes produced slower and less accurate responses than late probes because they appeared at the beginning of the stories and were therefore less accessible.

RTs to probe

Table 2.4 Experiment 4: Mean RTs to probe of state stories in milliseconds

	Early Probe		Late Probe	
	Mean (ms)	SD	Mean (ms)	SD
Long	1422.78	382.82	1247.67	305.90
Short	1471.20	442.92	1250.67	318.14

Repeated ANOVAs were carried out with RTs to probe as the dependent variable and duration (long, short) and probe position (early, late) as independent variables, and these showed the main effect of probe position ($F_1(1,67) = 49.61$, $p < 0.001$; $F_2(1,23) = 7.48$, $p = 0.012$). No main effects of duration or interaction were found. Means and standard deviations of the reaction times to early probe and late probe in the long and short condition are presented in Table 2.4.

2.5.3 Discussion

In this state experiment, only the probe position effects were found and no duration effect was observed, which was inconsistent with the previous results for event stories. I speculate that this might to be due to the fact that states are typically persistent and stable over time (there are no changes referred to). Therefore, with a longer or shorter interval, the activated information relating to a state did not necessarily change in the same way as was found for an event. In addition, unlike the event stories in which the actions referred to were more complex with longer durations, the interpretations of the state stories did not alter as a function of duration.

Null results are difficult to interpret. It can be argued that the change in the story structure made a difference, as the temporal adverbials appeared at the end of the third sentence in the present experiment. Nevertheless, this is unlikely to be the reason. As suggested by situation model accounts, when people read a text, they construct a mental model or a situation model for the described events, and when they are required to retrieve information about the events, instead of retrieving it from their memory of text, they retrieve it from the situation model of the event (Zwaan & Radvansky, 1998). Therefore, the positions of temporal adverbials should not significantly affect memory recall, as duration information should have been integrated into the situation model. Nevertheless, it is possible that there was not sufficient statistical power to enable an effect to be observed. Given that the construction of the items made it quite difficult to obtain plausible stories, fewer items were tested in the present experiment, compared to experiment 1.

2.6 Summary of the results and general discussion

In this chapter, I investigated how people process duration information of singular events conveyed by temporal adverbials and the underlying mechanisms of duration representation in language comprehension. There were two principal findings: first, in Experiments 1 and 2, it was shown that longer events took longer to process. Specifically, in Experiment 1, we observed the duration effects in processing cost for large-timescale events (e.g., it took two weeks to build a wall VS. it took three months to build a wall). In Experiment 2, we obtained a timescale effects (events with durations of days, weeks or months VS. events with durations of hours or minutes) in processing time. Moreover, we also explored underlying mechanisms for event representation in Experiment 3. We analysed the associations elicited by the large-scale stories and examined the variability of associated background knowledge by calculating the difference in proportions of event associations and state associations. The results showed that longer events tended to be associated with more diverse background knowledge (smaller event-state difference) and there was a correlation between diversity of activated background knowledge and processing costs. Specifically, the higher diversity in background knowledge, the slower response time to probes. The findings provide a possible explanation for the observed duration effects in Experiment 1 and suggest that activated background knowledge plays a role on modulating processing cost associated with event duration. Moreover, in Experiment 4, we also explored the duration representations of state stories by manipulating temporal adverbials. However, we did not find a duration effects in responses to probes. It is possible that the different causal structures in event and states or insufficient statistical power explain the absence of this effect.

These results provide some insights into how readers represent situation models when the duration of events is manipulated. The differences across time-scales ('drawing a tiger' as opposed to 'building a bridge') can probably be explained in terms of the different schemas associated with the events and the verbs of the story, along the lines of previous results (Fecica & O'Neill, 2010; Lindsay *et al*., 2011; Coll-Florit & Gennari, 2011). The semantic knowledge associated with a verb gives clues as to the duration of the events based on the different steps that might be involved in carrying out the event. These differences then lead to different retrieval times, as different knowledge is activated by the verbs. However, duration differences in processing time within large-timescale events are clearly not due to schematic knowledge but to the role of the specific temporal modification. The verbs and the stories were exactly the same across conditions except for the temporal adverbial. Yet, it was found that the processing cost associated with retrieving the long and short duration differed; longer events took longer to process. This difference in processing time was explained by the differences in the knowledge associated with each story, despite the fact that the lexical content was identical except for the indicated duration (one week, two weeks). These results were in line with previous findings on event duration processing in language, which argued that the background knowledge brought to mind by short and long events explains the differences in processing time observed (Coll-Florit & Gennari, 2011).

This type of explanation differs from embodied accounts, which argue for some analogical simulation of the event duration: whereas simulation proposals only provide a new concept to describe the phenomenon (long events are simulated for longer), knowledge-based accounts provide a mechanistic explanation rooted in schemas and semantic theories. In particular, this account explains retrieval cost as a function of the variability of the associated knowledge activated. Variability or diversity in associated knowledge across different semantic domains makes words harder to retrieve because the individual strength of each association is relatively weak and possibly different associations might compete with each other. This explanation is consistent with independent accounts of concrete and abstract concepts (De Groot, 1989) as well as with neural network accounts of ambiguity resolution (MacDonald *et al.*, 1994). Thus, the diversity account attempts to provide a possible mechanism for why long durations take longer to process, whereas embodied accounts do not specify the nature of the simulation, or what processes are engaged in simulations only lasting a few milliseconds during processing. Therefore, although more research needs to be conducted to support the diversity account, the present data suggest that semantic diversity accounts can be extended to explain effects resulting from temporal adverbial modification, thus going beyond duration effects resulting from lexical meaning or schemas.

The results are also consistent with segmentation accounts of event representations. Although the number of sub-events involved in a long event might not in themselves account for processing cost, the fact that longer events have more sub-events might give rise to more diverse associations. For example, although we know that building a bridge will entail many sub-actions, the duration of this event made explicit in the story (months or weeks) might be a cue to retrieve more possible sub-events as well as further related things (such as construction delays), thus making the associated knowledge more diverse. The temporal adverbial might also be a cue to imagine different kinds of bridges, bigger or smaller, and therefore again lead to different situation models being activated in the story. Although the relationship between the number of sub-events and the diversity of associations remains to be investigated, these two properties of event representations are not necessarily contradictory and may be linked to each other.

More generally, if the present data can be generalized further, it would be possible to give a unified account of the representation of event durations in language comprehension, regardless of whether the duration is elicited by verb semantics or temporal adverbials. This is important for understanding how the mind represents time more generally, for example, in relation to the events taking place and the potential events which could have taken place in a given interval according to general world knowledge. It is the human ability to activate likely events and make predictions according to previous schemas that underlies the associations elicited by stories and words.

2.7 From language to experienced events

As human beings, to understand how long an event is (e.g., how long does it take to drive from London to York?), we can either google the question or ask a friend. That's to say, we would get the answer through either written or spoken language. Alternatively, we could also answer the question about event duration through our own real-world experience (e.g., I can estimate how long it took me to drive to York from London last year). In this chapter, we have explored how people represent explicit event duration in language. We found that longer events took longer to retrieve, as longer events were associated with more diverse or more variable background knowledge. Then, how do we obtain temporal duration of a past event via our own experience? To have a better understanding of how we retrieve duration information of a past event through real-word experience. In the next chapter, I will report my published paper, which focuses on the representations of event duration in episodic memory (Wang & Gennari, 2019).

CHAPTER 3

How language and event recall can shape memory for time

DECLARATION:

The study presented in Chapter 3 is a published paper (see reference listed below), which contains original work completed by the author under the supervision of Dr. Silvia Gennari. Both authors contributed to the writing. Yaqi Wang analyzed data and wrote up the first draft of the paper.

Wang, Y., & Gennari, S. P. (2019). How language and event recall can shape memory for time. Cognitive psychology, 108, 1-21.

3.1 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.

3.2 Overview

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, 1985). 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.

3.2.1 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.

3.2.2 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, et.al, 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 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. To be more specific, the linguistic concepts entails slower or faster speed would translate into longer replay or shorter replay, because people tend to simulate the indicated speed and replay it as happening at a different speed. Therefore, while replay the temporal development of the event, they reconstruct the whole event, which should be different as a function of language.

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.

3.2.3 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 3.1. 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 3.1) 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.

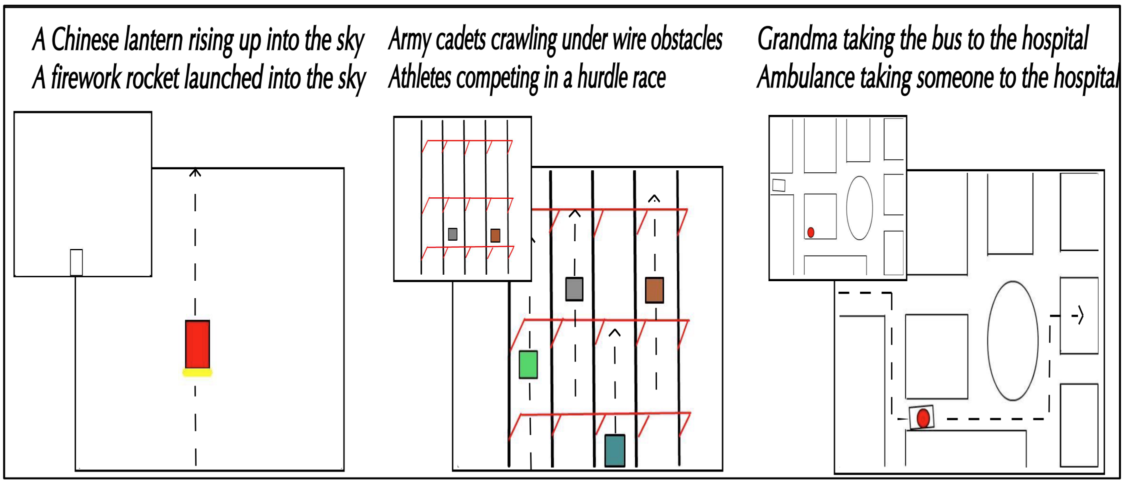


Figure 3.1 Current studies: Examples of the experimental stimuli used in Experiments 1 to 4. Descriptive phrases are shown on the top. The small squares show a still frame near the beginning of the animation (Figure 3.1 1the cue frame). The large squares represent the animations, where the arrows indicate 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.

3.3 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 3.1, 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.

3.3.1 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 $\alpha = 0.05$ 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 3.1). 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 3.1). 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 3.1 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 ($t(40) = -6.82$, $p < 0.001$).

Table 3.1 Experiment 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 segmentation is 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 =< 0.0001) (see Supplemental Materials for details). Table 3.2 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 3.2 Experiment 1: 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 3.2 for details of the trial structure. No mention of time was made at this stage.

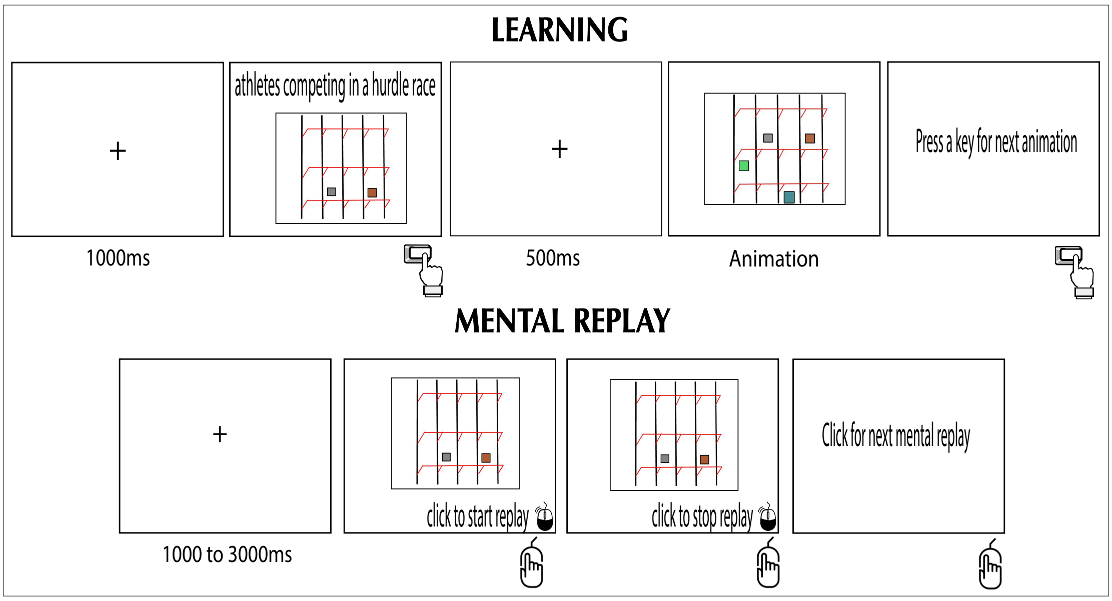


Figure 3.2 Experiment1: Schematic representations of the learning and replay tasks in Experiment 1.

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 3.2). 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. Deviation index is proportional to the total duration of an event, so the factors that go into the deviation index models should also be proportional to duration or number of segments (e.g. words per segment).

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.

3.3.2 Results

Reproduced Duration

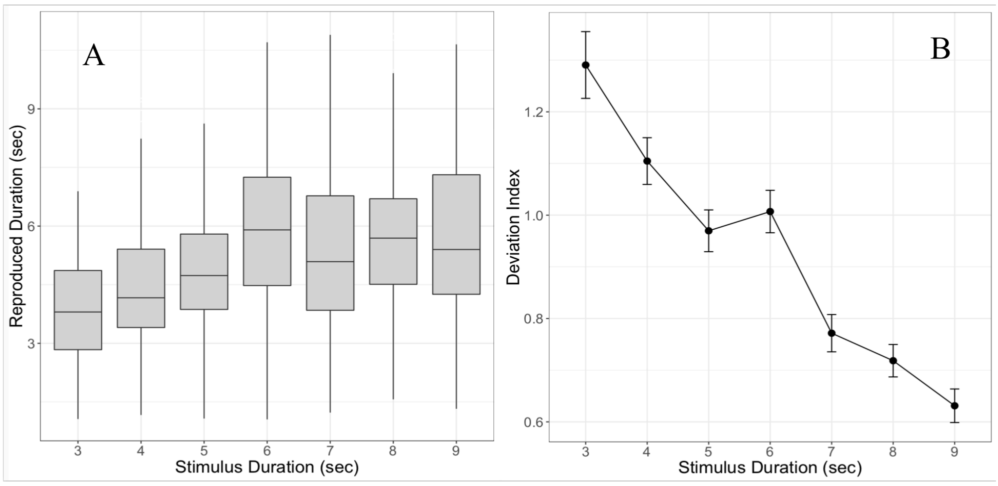


Figure 3.3 Experiment1：Results of Experiment 1. Panel A shows the medians and interquartile ranges of the reproduced durations as a function of the stimulus duration. Panel B shows the deviation index (reproduced duration/stimulus duration) as a function of the stimulus duration. Error bars indicate standard errors calculated according to Loftus and Masson (1994).

Table 3.3 Experiment 1: Modeling results for Experiment 1A

Dependent Variable	Fixed effects	Estimated coefficient	Standard error	t-value
Reproduced Duration (ms)	(Intercept)	2247.10	440.66	5.10
	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 Index	(Intercept)	0.45	0.09	4.81
	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 ($p < .05$).

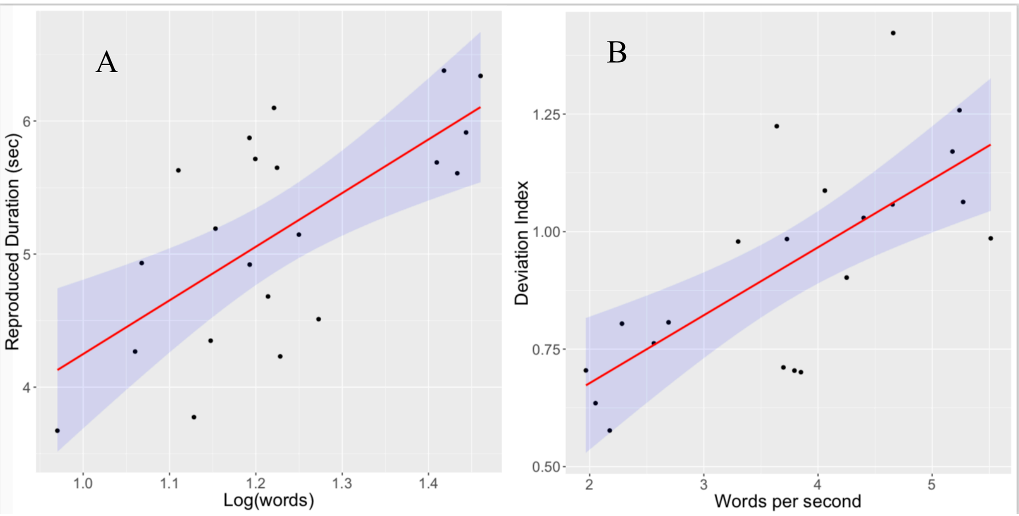


Figure 3.4 Experiment 1: Panel A shows the relationship between the mean reproduced duration and the mean log-transformed number of words recalled for each of the stimulus animations. Panel B shows the relationship between the mean deviation index (reproduced duration/stimulus duration) and the mean number of words per second (number of words recalled/stimulus duration) for each stimulus animation.

Mixed-effects modeling assessing the effect of language on reproduced duration revealed no significant effect of language ($\chi^2(1) = 0.09$, $p = 0.75$), 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 ($\chi^2(1) = 0.30$, $p = 0.58$). Figure 3.3 A 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 3.3 B 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 3.3 A, 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 3.3 B, 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 3.3.

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 ($\chi^2(1) = 5.75$, $p = 0.02$): the more information recalled, the longer the reproduced duration (Table 3.3). 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 ($\chi^2(1) = 9.99$, $p = 0.002$): 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 3.4A shows the relationship between the mean reproduced duration and the log-transformed number of words for each of the 21 stimulus animations. Figure 3.4B shows the relationship between the deviation index and the number of words recalled per seconds in the 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 3.1).

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.

3.3.3 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 3.3 A). 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 3.4B) 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.

3.4 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 3.4B), and therefore, there is more room for memory improvements in these cases. The effect of exposure should thus not be the same across all stimulus 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.

3.4.1Methods

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.

3.4.2 Results

Reproduced duration

Table 3.4 Experiment 2: Models' summaries for the results of Experiment 2

Dependent Variable	Fixed effects	Estimated coefficient	Standard error	t-value
Reproduced Duration (ms)	(Intercept)	1201.65	354.08	3.39
	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 Index	(Intercept)	0.53	0.06	8.57
	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).

Model comparisons assessing the effect of language indicated that there was no language effect in reproduced durations ($\chi^2(1) = 0.08$, $p = 0.93$) or deviation indices ($\chi^2(1) = 0.07$, $p = 0.78$). 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 ($\chi^2(1) = 5.38$, $p = 0.02$), 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 ($\chi^2(1) = 14.72$, $p = 0.0001$). 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.

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: $Z = -1.00$, $p > 0.05$).

Table 3.5 Experiment 2: Mean number of words produced in the recall task for Experiments 1 and 2 as a function of stimulus duration

Stimulus Duration (sec)	Experiment 1 Mean	(SD)	Experiment 2 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

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 ($\chi^2(1) = 8.23$, $p = 0.004$), 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 3.6 Experiment 2: Model summaries for comparisons between Experiments 1 and 2

Dependent Variable	Fixed effects	Estimated coefficient	Standard error	t-value
Reproduced Duration (ms)	(Intercept)	3798	380	9.99
	Stimulus duration	**204**	**69**	**2.95**
	Exposure	**-481**	**116**	**-4.14**
	Exposure × stimulus duration	**103**	**24**	**4.28**
Deviation Index	(Intercept)	1.64	0.09	17.55
	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).

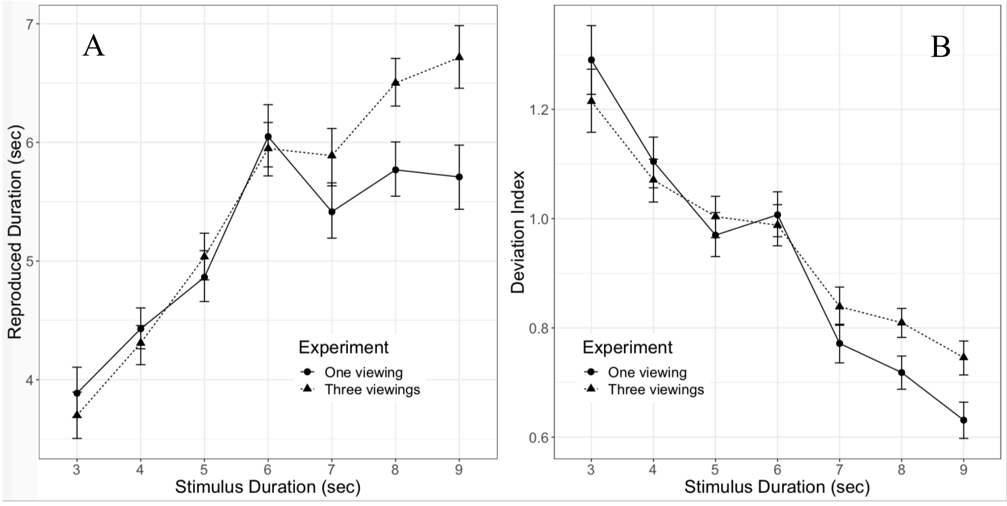


Figure 3.5 Experiment 2: Results of Experiments 1 and 2. The panels show the reproduced duration (in A) and the deviation index (in B) as a function of the stimulus duration. Error bars indicate standard errors calculated as in Loftus and Masson (1994) for each group.

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 ($\chi^2(1) = 16.99$, $p < 0.001$) and deviation indices ($\chi^2(1) = 12.62$, $p < 0.002$). The models are summarized in Table 6. Figure 3.5 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.

3.4.3 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 3.3A. 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.

3.5 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.

3.5.1 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

3.2, 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.

3.5.2 Results

Reproduced Duration

Table 3.7 Experiment 3: *Model summaries for the results of Experiment 3*

Dependent Variable	Fixed effects	Estimated coefficient	Standard error	t-value
Reproduced Duration (ms)	(Intercept)	2096.53	503.38	4.17
	Stimulus 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**
Deviation Index	(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).

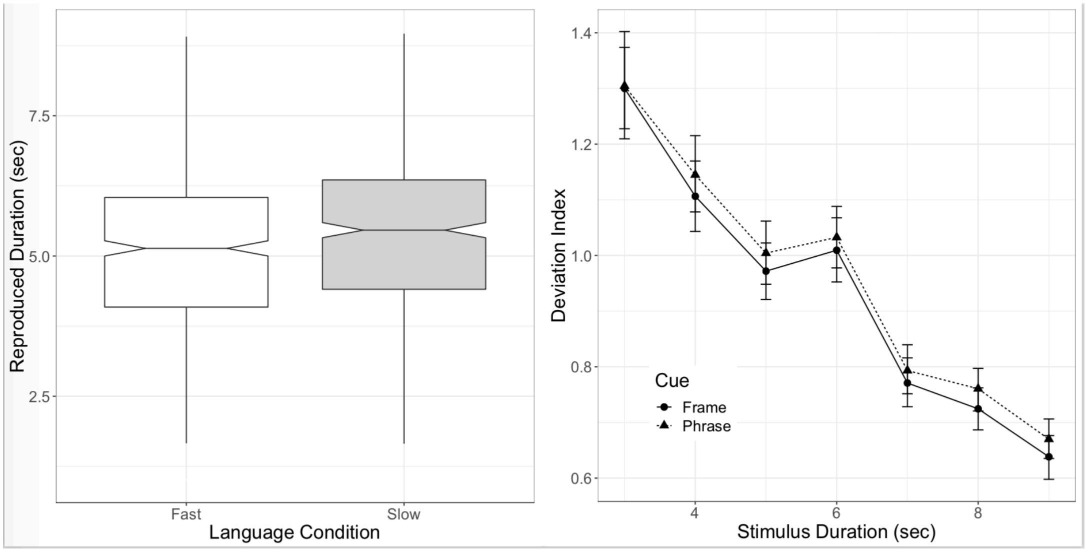


Figure 3.6 Experiment 3: Results of Experiment 3. Panel A shows the reproduced duration as a function of language condition. Notches indicate 95% confidence intervals. Panel B shows the deviation index collapsed across language conditions as a function of stimulus duration and cue type, i.e., Experiments 1 and 3. Error bars indicate standard errors calculated as in Loftus and Masson (1994) for each group.

Model comparisons revealed significant effects of Language for the reproduced duration ($\chi^2(1) = 8.05$, $p = 0.005$) and the deviation index ($\chi^2(1) = 8.21$, $p = 0.004$), 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 3.6A. 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 3.6B) 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 ($\chi^2(1) = 15.79$, $p < 0.001$) and deviation indices ($\chi^2(1) = 9.52$, $p = 0.002$), 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 3.7 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: $\chi^2(1) = 0.11$, $p > 0.05$; ratio: $\chi^2(1) = 0.88$, $p > .05$). As shown in Figure 3.6B, 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.

3.5.3 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.

3.6 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.

3.6.1 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%).

3.6.2 Results

Table 3.8 Experiment 4: *Model summaries for Results of Experiment 4*

Variable	Fixed effects	Estimated coefficient	Standard error	t-value
Reproduced Duration (ms)	(Intercept)	1985.02	562.91	3.53
	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 Index	(Intercept)	0.63	0.08	8.32
	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).

Statistical comparisons assessing the effect of language indicated a significant main effect for the reproduced duration ($\chi^2(1) = 10.03$, $p = 0.001$) and the deviation index ($\chi^2(1) = 9.16$, $p = 0.002$). 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 3.7A). 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 ($\chi^2(1) = 5.73$, $p = 0.02$) and deviation indices ($\chi^2(1) = 9.96$, $p = 0.002$), over and above the contribution of stimulus duration and number of segments (see Table 3.8). 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.

Comparisons across experiments 3 and 4: exposure effect

Table 3.9 Experiment 4: 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 3.10 Experiment 4: Model summaries for comparisons across Experiments 3 and 4

Variable	Fixed effects	Estimated coefficient	Standard error	t-value
Reproduced Duration (ms)	(Intercept)	3378	615	5.49
	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).

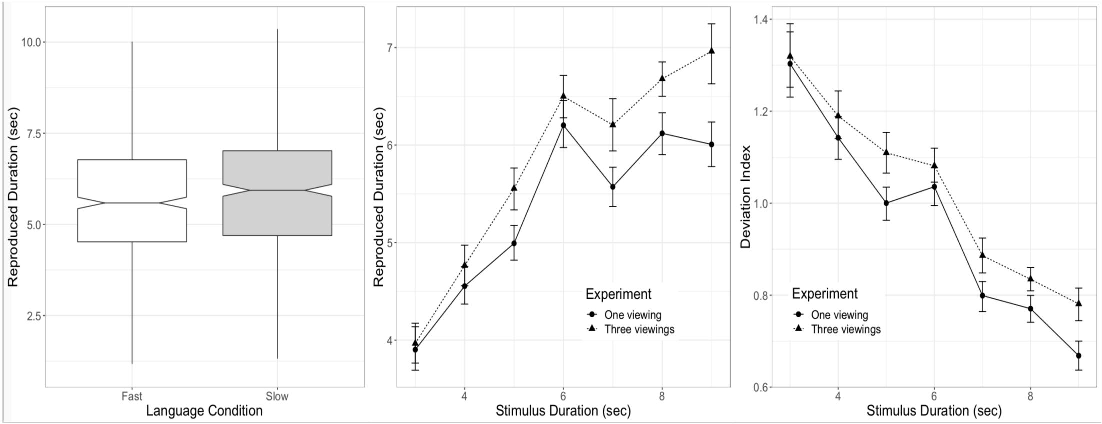


Figure 3.7 Experiment 4: Results of Experiment 4. Panel A shows the reproduced duration as a function of the language conditions. Notches show confidence intervals. Individual scores were adjusted according to Loftus and Mason (1994). Panels B and C show the reproduced duration and the deviation index 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.

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 (χ^2(1)= 9.85, p=0.002) and no interaction with stimulus duration (see Table 3.9). 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 (χ^2(2)= 8.14, p=0.004) but only a main effect of exposure for the deviation index (χ^2(2)= 7.85, p=0.005), see Table 3.10 and Figures 3.7B and 3.7C. 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 3.7C. Visual inspection of Figure 3.7 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.

3.6.3 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.

3.7 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 3.1). 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 3.4), 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 3.2, 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.

CHAPTER 4

Language influences on event memory are mediated by consolidation across sleep

DECLARATION:

Chapter 4 is a publication, which has been sent to Journal of Experimental Psychology: General and is under review now. I conducted the experiments on my own with the collaboration of Professor Gareth Gaskell under the supervision of Dr Silvia Gennari's. Both authors contributed to the writing. Yaqi Wang analyzed data and wrote up the first draft of the paper.

4.1 Abstract

Can a commentator’s descriptions of concurrent events modulate how observers later recollect the events? Would recollection change after observers slept? Here we investigate these questions and specifically ask whether linguistic encoding of events may shape recollection before or after sleep. Participants studied cartoon-like animations that were described by phrases implying either slow or fast motion (e.g. *a Chinese lantern raising up, a firework rocket being launched*). They later replayed the animations in their minds, and verbally recalled the events. We found that mental replays were longer for slow than fast descriptions after sleep but not before, suggesting that event memories were distorted towards the features highlighted by the language. Verbal recall was also longer after sleep and included more combinations of visual and linguistic stimulus information. Replicated results suggest that language modulates event memory through the integration of episodic details and verbal concepts operating in sleep-dependent memory consolidation.

In the previous chapter we investigated the underlying mechanisms of reproduced event duration and how the process is modulated by language. We tested memory for event duration immediately after learning and manipulated the retrieval cues (visual cue VS. linguistic cue) to explore when linguistic information shapes memory for event duration. It was found that reproduced duration is predicted by the amount of recalled information. In addition, language effects only occurred when duration reproduction was cued by linguistic concepts. When the linguistic concepts were only presented at encoding stage, we did not observe the language effects. The results suggested that the integration of linguistic concepts and event memory occurs at retrieval stage.

However, as indicated in Chapter 1, memory for event is dynamic, and changes over time (Moscovitch, Cabeza, Winocur, & Nadel, 2016, Bartlett, 1932). Then how does memory for event duration changes over time? As reviewed in Chapter 1, memories for an event initially retains highly detailed perceptual information. But over time, fewer contextual details, the gist of original event, are retained in memory (Moscovitch, Cabeza, Winocur, & Nadel, 2016; Bartlett, 1932). Then, how does memory for event duration change with memory decay or memory stabilization? If only gist is retained after a period of memory consolidation, according to our recall-based account, we would expect to observe shorter reproduced duration, as episodic details would be forgotten.

Moreover, sleep has been shown to play a key role in memory consolidation (Ashton, Jefferies, & Gaskell, 2018; Payne et al., 2008; van der Helm et al., 2011; Rasch & Born, 2007; Smith, 1995; Stickgold, 2005; Walker & Stickgold, 2006). During sleep, memory becomes schematized or conceptualized. Would the linguistic concept be integrated with event details during sleep? If sleep-induced memory consolidation indeed occurs, we would expect to observe the language effects even when the linguistic concepts are absent at retrieval stage.

4.2 Overview

Suppose you saw a person exercising in the neighbourhood park when your companion says *look, that's Jane running.* Will this event description influence how you later remember this event compared to other descriptions such as *walking or jogging*? Would this influence be any different if you recall the event before or after sleep? Here we investigate these questions by examining how people mentally replay and verbally recall past events that were conceptualized through language. We specifically ask whether linguistic descriptions at encoding may distort event recollection after a delay in which participants either slept or remained awake. This issue is critical in order to understand the role of language on event memory transformations over time, and more generally, the relationship between language and memory.

An enduring debate in language and memory research concerns whether and how verbal encoding may shape memory representations. Interactive-encoding accounts propose systematic language influences at encoding: Labelling scenes or objects activates conceptual features that distort the encoded representations towards the features highlighted by the labels, resulting in failures to recognize or retrieve previously seen stimuli (Carmichael, Hogan, & Walter, 1932; Feist & Gentner, 2007; Lupyan, 2008, 2012). Retrieval-based accounts in contrast argue that labels and semantic schemas more generally, may covertly operate at retrieval rather than encoding because they provide useful retrieval cues, particularly in demanding memory tasks. Thus, labels need not systematically shape the encoded representation (Alba & Hasher, 1983; Hanawalt, 1937; Hanawalt & Demarest, 1939; Neuschatz, Lampinen, Preston, Hawkins, & Toglia, 2002). This prior research however has mostly investigated the role of alternative labels in object and scene memory. The few studies examining event memory have targeted systematic and ingrained behavioural differences across speakers of different languages, e.g., (Gennari, Sloman, Malt, & Fitch, 2002; Papafragou, Hulbert, & Trueswell, 2008; Trueswell & Papafragou, 2010), and thus are not necessarily comparable to those examining labelling differences within a language. Nevertheless, these cross-linguistic studies suggest no influence of event descriptions on immediate memory recognition, despite language-related differences in other tasks. Thus, unlike objects' labels, verbal encoding of events might not modulate immediate memory representations at encoding or retrieval.

An intriguing possibility that has not been investigated, however, is that language may systematically influence event memory representations, but only after a period of memory consolidation. Sleep-dependent memory consolidation in particular is known to integrate or combine new episodes with prior knowledge in semantic memory (Landmann et al., 2014; Moscovitch, Cabeza, Winocur, & Nadel, 2016; Rasch & Born, 2013; Walker & Stickgold, 2010). Newly learned words, for instance, compete during recognition with existing words after 12 hrs of sleep but not of wakefulness (Dumay & Gaskell, 2007; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). Since verbal descriptions at encoding evoke familiar event concepts or schemas, episodic event features might become integrated with these verbal concepts after sleep. This possibility therefore predicts that an event that has been construed as either *running* or *walking* at encoding should be mentally replayed as faster or slower respectively after sleep but not before, thus suggesting that the event features conveyed by the language have shaped the event memory representation during consolidation. This influence should be specifically observed when visual cues, rather than verbal ones, prompt event recollection so that language cannot overtly influence retrieval processes, as previously demonstrated (Burt & Popple, 1996; E. F. Loftus & Palmer, 1974; Wang & Gennari, 2019).

An interesting additional question about mental replays is whether they become on average shorter or longer after sleep regardless of the specific verbal descriptions. If sleep promotes integration of episodic and conceptual features, as argued by the integration account above, mental replays from memory may become overall longer after sleep relative to wakefulness because the episodic representations underpinning replays should be augmented with conceptual features. This expectation is consistent with prior research showing that memory-based stimulus replays and time estimations vary as a function of the amount of information recalled: more segments perceived in the stimuli and more stimulus details recalled lead to longer stimulus reproductions and duration estimates (Burt & Kemp, 1991; Faber & Gennari, 2015; Ornstein, 1969). In particular, the longer the verbal recollection of events, as measured by the number of words used, the longer their mental replays (Wang & Gennari, 2019). These results therefore suggest that detail-rich event memories tend to be reproduced as longer, and therefore memory enrichment during sleep might be reflected in longer mental replays. Note, though, that sleep is also associated with a variety of memory transformations such as gist extraction or multi-item generalization (Landmann et al., 2014; Payne et al., 2009; Stickgold & Walker, 2013), and these transformations potentially entail different outcomes. Gist extraction, in particular, might reduce the level of episodic detail associated with the events, leading to shorter mental replays after sleep.

Under the integration account, the linking of episodic and verbal conceptual features during sleep may also result in different verbal recall performance, in particular, verbal recall might be longer and composite in nature. When communicating everything they can remember about the stimuli, participants should tend to combine verbal concepts with perceptual features more often after sleep than before due to the integration mechanism operating during sleep. This possibility therefore predicts that the number of words used in verbal recall should increase after sleep, and words belonging to the linguistic descriptions should be more readily combined with those describing perceptual features. Again, though, if sleep's primary effect in these circumstances is gist extraction, then one might expect the opposite pattern of results.

The present study tested these predictions. Participants first studied cartoon-like animations varying from 3-9s. Each animation was described by a phrase implying either a fast or slow motion, e.g. *a Chinese lantern raising up, a firework rocket being launched*. The phrases provided clues to understand the animations, which were unspecific as to the nature of the moving object (Figure 4.1). Two groups of participants were tested 12 hrs after learning with or without intervening sleep (Wake and Sleep groups). We compared these data to a group tested 10 minutes after encoding from a previous study that used identical methods (Wang & Gennari, 2019) (Reported in Chapter 3, Experient 4). At test, participants were asked to mentally replay or reproduce the animations exactly as they occurred in their original time course and to verbally recall everything they could remember about the animations when cued by animations' frames (Figure 4.1). We measured the duration of the mental reproduction, and the number and type of words used in verbal recall.

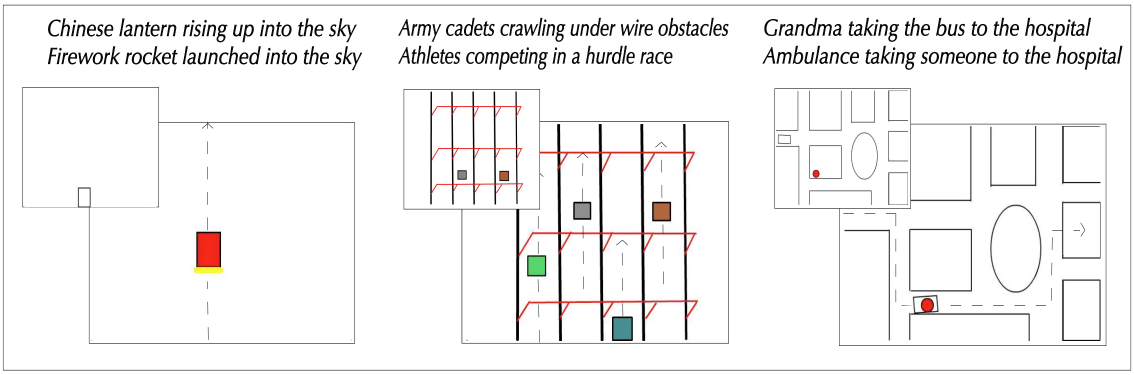


Figure 4.1 Example stimuli: Smaller rectangles represent the cue-frame. Larger rectangles represent the animations. Arrows indicate motion path.

4.3 Methods

Participants

134 English native speakers were recruited from the University of York (Sleep and Wake groups). They participated for course credit or payment. The study was approved by the Ethics committee of the Department of Psychology, University of York, UK. Previous sleep studies have shown sleep-dependent integration effects with 32 participants per group and a single stimulus list (Dumay & Gaskell, 2007). Since our design had two stimulus lists, we aimed to test 33 or 34 participants per list because we expected having to exclude participants, as in previous studies. From the total 134 participants, 8 participants were excluded because more than 50% of their event reproductions fell within the lowest 5% or the highest 95% of the observations. This was expected because the mental replay task was relatively unconstrained (see below), and participants could simply click through the trials without mentally reproducing the stimuli. Thus, participants who had many reproductions below 1s or over 15s did not follow

instructions or misunderstood the task. The total number of participants for the Sleep group was 62 (31 per list) and for the Wake group was 64 (32 per list). The Immediate group that was used as a comparison contained 52 participants previously tested (as reported in Wang & Gennari, 2019, experiment 2) plus 10 additional participants, who were tested in the present study to match the size of the Sleep and Wake groups.

Stimuli

21 cartoon-like animations varying from 3-9s in increments of 1 sec were created using Adobe Flash software. Animations consisted of simple geometric figures moving (e.g., squares or circles) within a familiar setting. There were three animations per duration. Each animation was paired with two descriptive labels providing information about who or what the geometric figure was meant to represent (see Figure 4.1). The description thus indicated what type of object the figure represented and implied different motion speeds (slow- or fast-moving entities). In seven animations, the speed information was conveyed by the verb (e.g., *children walking vs. running to their classrooms*). Example animations can be found at https://sites.google.com/york.ac.uk/stimuli/. Each animation was also paired with a frame near the beginning of the animation that was later used as a cue to recollect the corresponding animation (see Figure 4.1). The cue frames and accompanying phrases for all animations are listed in the Supplemental Materials. 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, as 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, e.g., a street scene closer or farther away 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 (see Table 4.1). All ratings studies showed no significant difference across conditions (all p's > .10), except for the implied speed rating ($t(40) = -6.82$, $p < 0.001$).

Table 4.1 : Mean and standard deviations from pre-test studies

Rating Tests	Condition	
	Slow	Fast
Phrase implied 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.

Design and procedure

All groups performed identical learning and memory tasks, but the memory tasks took place at different time points after learning. The immediate group (as reported in Chapter 3, Experient 4) performed the memory tasks 10 minutes after learning and a distractor task (math calculations) . The immediate group (as reported in Chapter 3, Experient 4) was primarily tested in the afternoon (35% of participants in this group were tested between noon and 2PM, 47% were tested between 2PM and 4.30PM, and 18% were tested before noon). The other two groups were tested twelve hours after learning, but they differed in whether sleep took place during this period. The Wake group learned the animations in the morning (e.g., 8AM) and was then tested 12 hrs later in the evening (e.g., 8PM). The sleep group in contrast learned the stimuli in the evening (e.g., 8PM) and was then tested 12 hrs later in the morning of the following day (e.g., 8AM). See Figure 4.2. Each animation-phrase pair was assigned to one of two possible stimulus lists so that each participant was exposed to all 21 animations but only in one language condition (10 in one condition, 11 in the other). Half of the participants in each group were randomly allocated to each list.

At the end of the experiment, the Sleep and Wake groups completed the Pittsburgh Sleep Quality Index questionnaire (PSQI) (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989), the Stanford Sleepiness Scale (SSS) (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) and the Morningness-Eveningness Questionnaire (MEQ-SA) (Horne & Ostberg, 1976). From these questionnaires, we checked that participants in the Wake and Sleep groups did not differ in their sleepiness or alertness at the time of test, the quality of their sleep and their alertness at different times of the day. There was indeed no difference across groups in these measures. The mean rating for the SSS were 2.5 for the Wake group and 2.8 for the Sleep group ($t(124) = -1.79, p > 0.05$). The means for PSQI were 6.2 and 5.9 respectively ($t(124) = 0.834, p > 0.05$) and for the MEQ-SA 9.3 and 8.4 ($t(124) = -1.29$, $p > .05$). These findings rule out the potential influence of alertness or chronotype between participants due to the time of testing.

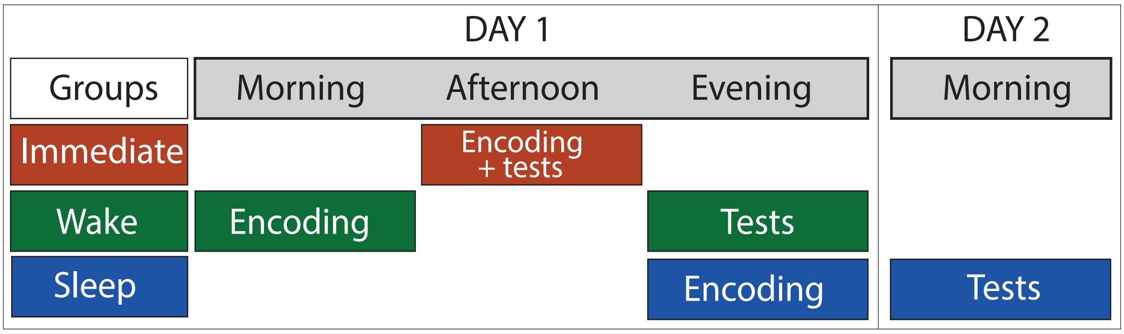


Figure 4.2 Schematic representation of the study sessions in each group

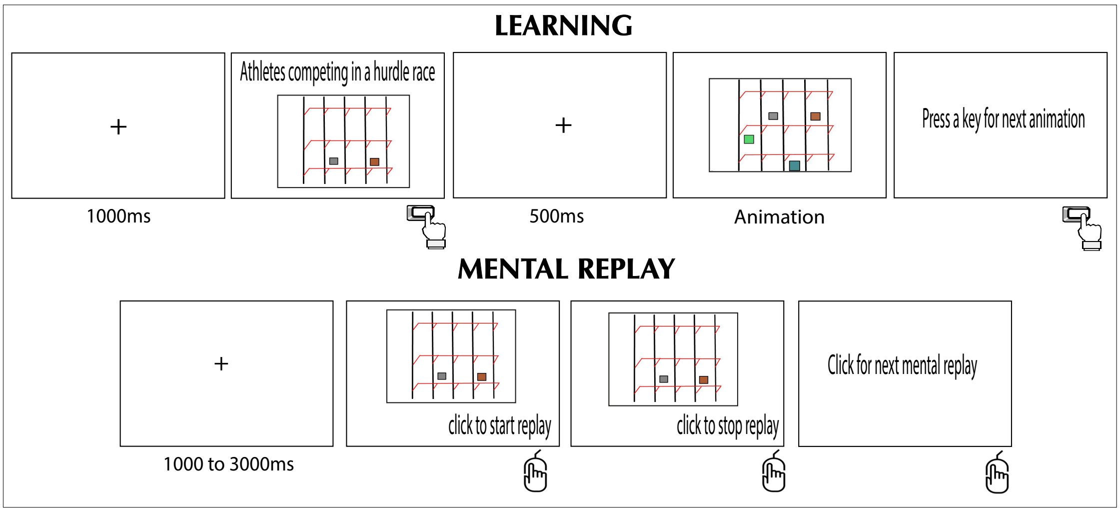


Figure 4.3: Trial structure in the learning and mental replay task

Stimulus Learning. Participants were instructed to study the animations for a subsequent memory test. They were told that the phrases provided clues 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 3). After participants had cycled through all the animations, they were told that they would watch them twice more to make sure that they learn them well. The programme then cycled through the stimuli in random order with an intervening screen announcing additional viewings. There was no mention of the upcoming mental replay task at this stage.

Memory tasks. The memory tasks included the *mental replay task* and *verbal recall task* in this order. In the mental replay task, participants were told to mentally replay the animation indicated by the cue-frame (Figure 4.1) exactly as it occurred in its original time course. We adopted this task because it involves mapping the events stored in episodic memory into a mental replay that must unfold over clock time and thus captures reproduced duration. This task is also similar in nature to duration reproductions and episodic simulations used in time and memory research respectively (Grondin, 2010; Schacter & Addis, 2007). Instructions provided examples of how the task would proceed and indicated that participants should replay the movie in their minds, exactly as they saw it. See trial structure in Figure 3. The duration between the Start click and the Stop click for each mental replay was considered as the reproduced duration for each animation. The order of presentation of the cue-frames was random. Participants had to keep their dominant hand on the mouse throughout the task for more accurate timing. After the mental replay task, participants were asked to type in 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 including physical characteristics (colour, shapes, etc.). Each trial started with the presentation of the cue-frames in random order and participants wrote their recollection in a textbox that appeared underneath the cue-frame. After this task, participants in the Wake and Sleep group filled out a questionnaire about their sleep and level of tiredness. For a replication of the Sleep manipulation, see Supplemental Materials.

Data treatment and statistical analyses

The main dependent measures for the mental replay tasks were the reproduced duration in milliseconds and the ratio between reproduced duration and actual clock duration typically used in time research. We refer to this ratio, expressed as a decimal fraction, as the deviation index as it indexes the extent to which reproduction duration deviates from the actual stimulus duration: a ratio of 1.0 indicates an accurate reproduction, whereas lower or higher values indicate under- or over-reproductions. For the analyses of reproduced duration and deviation indices, we excluded trials faster than 1s and those with incorrect verbal recall. Participants generally recollected details of the animations with 99% accuracy in all groups. The recollection of any number of correct details was considered as correct recollection. To minimize the influence of outliers in our analyses, we used the Tukey method for outlier removal. We first used the groups' interquartile ranges (IQR) to identify outliers above the third quartile + 1.5 × IQR. This resulted in values over 12.4s and 12s for the Sleep and Immediate groups (as reported in Chapter 3, Experient 4) and over 10.7s for the wake group being excluded. We used the same procedure to identify outliers within each individual's distribution of scores (i.e., scores higher than the third quartile + 1.5 × the IQR of an individual's distribution). Together these exclusions comprised less than 6% of the whole data set.

Hypotheses testing used linear mixed-effects models carried out in the *lme4* package of *R* (version 3.4) and lmerTest package (Bates, Maechler, Bolker, &

Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). The models included the maximal random effects structure allowed by convergence (i.e., random intercepts for subjects and items, by-subjects random slopes for stimulus duration and by-item random slopes for language condition) (Barr, Levy, Scheepers, & Tily, 2013). Stimulus duration was included in the models for reproduced durations because these measures are known to correlate with each other. Since deviation indices are computed from the stimulus duration, this factor was not included in the model for deviation indices. We used Helmert contrasts to specifically compare across groups: Immediate vs Delayed testing (i.e., Immediate vs Wake + Sleep groups coded as -2 1 1) and Sleep vs Wake group (coded as 0 -1 1). The fast and slow language conditions were coded 1 and -1. Post-hoc simpler contrasts to unpack interactions were conducted by comparing estimated marginal means with p-values adjusted using the Tukey method as specified in the *emmeans* package (https://CRAN.R-project.org/package=emmeans). Data and models are available at https://osf.io/k8yxa/?view_only=826cf0069db946a08b11d0cb3e728136.

For the verbal recall data, we compared the number of words produced, after removing introductory phrases such as *the title of this animation was …*. to examine the effect of sleep on verbal recall. Word counts can be seen as an objective index of the amount of information remembered about the animation and have been shown to strongly correlate with hand-coded number of details described (Wang & Gennari, 2019). Moreover, this measure has the advantage of being objective, as it does not depend on arbitrary definitions of what should count as a detail, and thus provides a handy approximation to how much people remember about an event. As shown in Table 4.2, most content words such as nouns, adjectives, adverbs and verbs indicate details of the animations, whereas function words such as articles, which add to the word count, are unlikely to systematically differ across groups. For statistical analyses of word counts, we followed the same procedure described for reproduced durations to remove outliers, which constituted 5% of the data set. Mixed-effects models were performed with word counts as dependent variable and group and duration as fixed factors (the random structure included random intercepts for items and subjects, random slopes for stimulus duration plus random slopes for groups). Identical models using the log-transformed word counts yielded a similar pattern of results.

To examine whether the combination of conceptual and perceptual words in verbal recall increased with Sleep, we examined the type of referring nouns used by participants. We reasoned that referring to objects as geometric shapes, e.g., squares and circles, indicates that participants brought to mind the visual characteristics of the animations, whereas referential nouns from the stimulus phrases, e.g., *athletes* or *bus*, indicated reliance on conceptual information absent in the animations. A programme automatically searched for the occurrence of pre-defined critical nouns (each object in the animations was referred to by a limited number of nouns, which were manually compiled from the whole data set). Based on this noun search, each trial was classified as containing *only phrase nouns* or *only shape nouns* depending on whether the nouns were in the stimulus phrases (e.g., *athletes* or *cadets* for the middle animation in Figure 4.1) or they refer to the shape of the entities (e.g., *square*, *rectangle*). Close synonyms of these nouns (e.g., *runners, soldiers, boxes, cubes*) were classified accordingly. If a trial contained at least one shape and phrase noun, it was classified as containing *mixed nouns*. The number of times a particular noun occurred in a recall trial did not change trial classification. Examples of this classification can be found in Table 4.2 for each of the three animations in Figure 4.1. The words in bold represent the words that were identified in the trial as belonging to the phrase noun category. Underlined words indicate those that were identified as shape nouns. The set of words that were searched in the text at available online alongside the data. This classification was then analysed with logistic mixed-effect models to test whether mixed trials occurred more often in the verbal recall of the Sleep group, compared to the Wake and the Immediate group (as reported in Chapter 3, Experient 4). The random structure in these models included random by-subjects and by-items intercepts. The number of unclassified trials was less than 1% in all groups.

Table 4.2 : Examples of coding for recall data

Animation recall	Phrase nouns	Shape nouns	Mixed nouns
*The **rocket** ignites and moves directly upwards to the top of the frame.*	1	0	0
***Rocket** shoots upwards, the rectangle changes from light orange to a red and there is a yellow line below. **Rocket** moves straight upwards.*	0	0	1
The pale rectangle turned yellow, orange and red and travelled upwards, with a red patch around the bottom edge.	0	1	0
*The grey **hurdler** starts off slightly faster than brown, but after the second **hurdler** brown overtakes and wins the race.*	1	0	0
*Two more squares appear as the grey and brown square **cadets** approach the finish. The brown **cadet** is the first to cross the finishing line, the other two **cadet** squares are blue and purple.*	0	0	1
The grey square finishes first. followed by brown. A lime green square appears, and a darker green square also appears which finishes last.	0	1	0
***Grandma** getting bus to **hospital**, goes to the top of rounded building on the way to middle building furthest left which turns orange.*	1	0	0
*The **ambulance** turns right, then left. Picks up red circle. Continues straight, turns left, goes around the oval from the top. Drops red circle in second square on the right. The box changes to orange.*	0	0	1
The white rectangle moves to the right then down to the red ball, the red ball enters white rectangles, then both move to the right and then up past the left of the oval.	0	1	0

Notes: nouns in bold indicate those identified in each trial as phrase nouns, and underlined nouns were identified as shape nouns. The codes in the final three columns identify the categorization of the description as a whole.

4.4 Results

Reproduced duration and deviation index

Table 4.3: Modelling results for reproduced duration deviation indices and number of words

DV	Fixed effects	Estimate	SE	df	t	p
Reproduced Duration	(Intercept)	3533.96	271.97	26	12.94	<0.001
	Language	27.52	27.23	18	1.01	0.33
	Imm. vs Delay	287.08	57.56	186	4.99	<0.001
	Wake vs Sleep	-336.30	99.18	186	-3.39	<0.001
	Stimulus duration	370.53	43.12	27	8.59	<0.001
	Language: Imm. vs Delay	14.18	15.71	3313	0.90	0.37
	Language: Wake vs Sleep	-57.83	27.18	3316	-2.13	0.03
Deviation Index	(Intercept)	1.03	0.06	24	18.35	<0.001
	Language	0.01	0.01	18	0.98	0.34
	Imm. vs Delay	0.00	0.01	182	0.29	0.77
	Wake vs. Sleep	-0.05	0.02	183	-3.04	0.00
	Language: Imm. vs Delay	0.00	0.00	3312	1.36	0.18
	Language: Wake vs Sleep	-0.01	0.01	3314	-2.36	0.02
Words	(Intercept)	18.07	4.20	24.70	4.31	<0.001
	Stimulus duration	1.72	0.65	22.84	2.65	0.01
	Imm. vs Delay	1.33	0.45	149.99	2.97	0.003
	Wake vs Sleep	-2.09	0.73	181.51	-2.88	0.004

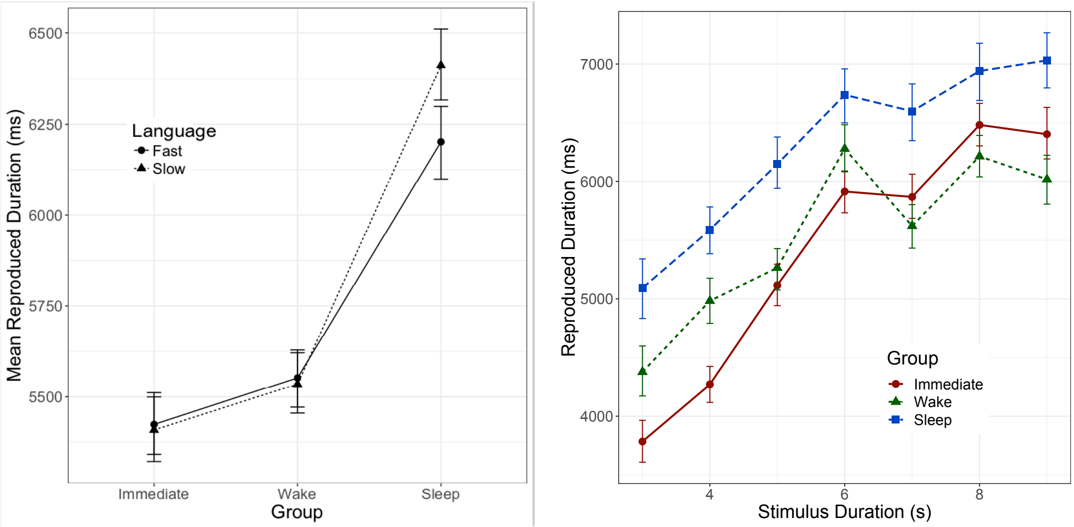


Figure 4.4: Mean reproduced event duration across participants as a function of group and language condition (left panel). Reproduced duration as a function of group and stimulus duration (right panel). Error bars were computed according to (G. R. Loftus & Masson, 1994)

We predicted that reproduced durations should lengthen overall, and also vary with language after sleep compared to wakefulness (i.e., a main effect of group and an interaction between group and language) because the conceptual features accompanying the animation should become integrated with the animations' episodic details overnight and should thus enrich the memory representation underpinning reproductions. A model testing these predictions revealed that both of these effects were significant, regardless of whether they were analyzed using reproduced durations or the deviation indices (see Table 4.3). As shown in Figure 4.4 (left panel), animations that were described as fast-moving at encoding were reproduced as shorter than those described as slow-moving after sleep but not immediately, or after wake. Deviation indices (reproduced duration/stimulus duration) showed a similar pattern to that of Figure 4.4, with larger indices after sleep indicating greater deviations from the actual stimulus duration (Mean and SD Immediate group: .95 (.35), Wake group: 1.0 (.50), Sleep group: 1.15 (.43)). Duration reproductions were on average accurate in the wake group, but 15% longer after sleep, and slightly underestimated immediately after encoding. Post-hoc contrasts between conditions per group confirmed a language effect in the sleep group for reproduced duration ($t = 2.35$, $p = 0.02$) and deviation indices ($t = 2.65$, $p = 0.009$) but not in the other group ($p > 0.05$).

For completeness, we also explored the relationship between stimulus duration and group, reported in detail in the Supplemental Materials. These models revealed that there was an interaction between the Immediate and the delayed-testing groups and stimulus duration, suggesting that the difference between these groups varied with the stimulus duration. As shown in Figure 4.4 (right panel), the Wake and Sleep group tended to lengthen short stimulus durations and shorten long ones, compared to the Immediate group. This tendency is commonly observed in time research (Jazayeri & Shadlen, 2010; Roy & Christenfeld, 2008; Wang & Gennari, 2019) and has been linked to the density of information recalled: the amount of information recalled per time units is proportionally larger for short stimuli compared to long ones (Roy & Christenfeld, 2008; Wang & Gennari, 2019). Thus, it appears that delayed testing exacerbates this memory bias, perhaps due to more uncertainty about the events' temporal development. However, there were no substantial differences in reproduced duration between the Wake and Sleep groups across stimulus durations (no significant interaction between group and stimulus duration), although there was a small one for deviation indices. This result suggests a relatively consistent increase in reproduced duration after sleep compared to wakefulness.

Taken together, these results suggest that event reproductions become longer after sleep and, critically, that the accompanying descriptions of the events shape how those events are recollected only after sleep-dependent consolidation has occurred. This important finding is novel and so we chose to try to replicate it. Since an additional experiment reported in Wang & Gennari (2019) has also found no language effect prior to sleep, we examined just the effect of language after sleep in our replication group. The replication is simply to check that the results are true. This successful replication is reported in the Supplemental Materials.

Amount and type of information recalle

We predicted that if sleep plays a role in augmenting episodic memories with the conceptual features presented at encoding, the amount of information recalled as measured by word counts should increase after sleep. Statistical analyses of word counts (as well as log-transformed counts) confirmed that the Sleep group produced more words in verbal recall than the Wake group ($t = -2.79$, $p = 0.02$) and the Immediate group ($t = 3.88$, $p = 0.0005$) but there was no difference between the Wake and the Immediate group in post-hoc comparisons, despite a numerical increase (Table 4.3, Figure 4.5). This result suggests that verbal recall lengthened after sleep. Because previous results have shown correlations between the amount of information recalled and reproduced duration for recently experienced events as well as distant ones (Burt & Kemp, 1991; Ornstein, 1969; Wang & Gennari, 2019), we checked whether this relationship also holds for our stimuli in each group. Word counts significantly predicted reproduced durations and deviation indices over and above the influence of stimulus duration, language and group. We compared the models in Table 4.3 to models additionally containing word counts and found a significant difference

($\chi^2(1) = 34.4$, $p <.0001$; $\chi^2(1) = 27.5$, $p < .0001$). These results indicate that the increase in reproduced durations and deviation indices across groups is linked to the increased amount of information recalled, suggesting that enriched event representations underpin both tasks, namely, mental replays and verbal recall.

We also predicted that if sleep promotes integration of episodic and verbal conceptual features presented at encoding, verbal concepts would be more readily combined with perceptual aspects of the animations after sleep than before. To test this possibility, we compared the number of trials in which Mixed Nouns were used, i.e., both Shape and Phrase nouns were alternatively used in the same recall trial to refer to the entities in the animation. Figure 4.5 shows the percentage of trials in which only Phrase, only Shape Nouns or Mixed Nouns were used in each group. A logistic model with Mixed Nouns coding as dependent variable revealed that there were main effects of groups when comparing Immediate vs. Delayed testing ($z = 2.99$, $p = .003$) and Wake vs. Sleep ($z = -2.07$, $p = .04$), indicating that mixed-nouns trials increased in frequency at delayed testing conditions taken together and after sleep. The difference between the immediate and delayed groups was carried by the Sleep group, as there was no significant different between the Immediate and Wake group in post-hoc contrasts ($t = -1.8$, $p = 0.18$), although there was a numerical increase in the Wake group relative to the Immediate group. There were no differences between groups when comparing Only Phrase or only Shape Nouns trials and no interactions. These results therefore suggest that after sleep, verbal recall becomes more mixed in nature, indicating a substantial increase in the combination of episodic and conceptual features of the animations.

4.5 Discussion

Does language modulate how we recollect events? Previous studies examining dynamic events cross-linguistically have failed to find influences of linguistic descriptions at encoding in subsequent memory tests (Gennari et al, 2002, Papafragou et al., 2008, Wang & Gennari, 2019), particularly when non-linguistic cues prompt recollection. The present results are the first to show that event descriptions at encoding can shape visually-cued event recollection. Specifically, event replays from memory were shortened or lengthened as a function of the speed of motion implied by the linguistic description, suggesting that event memories were distorted towards the features highlighted by the language. Importantly, this distortion was absent immediately after learning or after 12 hours of wakefulness and only emerged after sleep. This suggests that linguistic and visual event information remained stored as associated but distinct pieces of information before sleep, and they became unitized or integrated only after sleep. Thus, language influences on event memory are mediated by the integration of episodic and conceptual features operating during sleep.

Interestingly, regardless of the specific details of the linguistic description, event replays after sleep were consistently longer than those after wakefulness. This was observed in spite of the fact that a relationship to stimulus duration was retained after time spent asleep or awake (Fig. 4), suggesting a link to specific stimulus details. Replayed durations were also predicted by the length of verbal recall over and above the influence of stimulus duration. Consistent with previous findings (Burt & Kemp, 1991; Ornstein, 1969; Wang & Gennari, 2019), this result suggests that as the recalled information increases so does the reproduced duration. Nevertheless, there was a tendency for longer event reproductions and verbal recall in the Wake group compared to the Immediate group, suggesting that the integration of visual and linguistic information may start during wakefulness (Moscovitch et al., 2016; van Kesteren, Ruiter, Fernández, & Henson, 2012), but is accelerated and strengthened during sleep. Interestingly, the degree of integration that might have occurred before sleep was not strong enough for language to influence replayed duration or the type of verbal recall. Therefore, the concurrent influence of language, recalled information and stimulus duration on event mental replays provides strong evidence for the hypothesis that verbal concepts and episodic details are integrated during sleep, and specifically, that event information conveyed by the language augmented the event representation originally encoded.

Verbal recall results also indicated that after sleep, participants were more likely to include both referential nouns from the descriptions (e.g., *lantern*) and visual shape nouns (e.g., *square*) in the same trial, suggesting that the link between visual and conceptual information became stronger after sleep. At this point, the visual probe more strongly cued both the description and episodic details, resulting in more mixed trials. This result fits with previous findings showing overnight strengthening of learned associates (Landmann et al., 2014; Lau et al, 2010) and integration of episodic and schema features (Hennies, Lambon Ralph, Kempkes, Cousins, & Lewis, 2016; van Kesteren et al., 2012). It also converges with our findings above in suggesting that before sleep, perceptual and conceptual features were associated but kept somewhat distinct in memory, only becoming unitized and less dissociable after sleep.

The present results have implications for the relationship between language and event memory. Interactive-encoding accounts for scenes and object memory argue for systematic distortions at encoding in favour of the features highlighted by the language (Feist & Gentner, 2007; Lupyan, 2008; 2012). Retrieval-based proposals in contrast argue that language may be covertly recruited to aid retrieval or performance in difficult tasks (Alba & Hasher, 1983). Our results appear inconsistent with both these alternatives. The presence of linguistic influences on event replays only after sleep suggests that language did not shape event memories at encoding or retrieval before sleep. Moreover, the fact that the same task was used across groups and the same amount of time intervened before testing for the Wake and Sleep groups suggests that task difficulty was not the critical factor. Instead, our results suggest that verbal concepts at encoding modulated memory transformations during sleep and more generally, memory consolidation, resulting in distorted representations towards the conceptual features conveyed by the language.

The present findings also highlight the possibility that the role of language in event memory may differ from that of object labels, as in (Carmichael et al., 1932; Lupyan, 2008), although more within-language studies on event memory are needed. Unlike objects or scenes, dynamic events contain multiple features and dimensions (space, motion, actors, goals cf. (Kurby & Zacks, 2008)) that are less likely to necessitate discriminating labels at encoding or retrieval in naturalistic contexts. On the contrary, the presence of multiple episodic features, which are temporally and causally linked in a coherent sequence, may increase distinctiveness between events and thus may lead to episodic and verbal conceptual features being separately encoded before sleep.

In general, the present results also have implications for the role of language in event memory. Rather than promoting a different memory transformation during sleep such as gist extraction, verbal encoding was associated with integration of episodic and conceptual features, consistent with previous schema-based studies (Hennies et al., 2016). This role resembles that of prior knowledge or schemas in many memory consolidation studies (Moscovitch et al., 2016; van Kesteren et al., 2012) and suggest that verbal encoding may operate in similar ways. However, by highlighting different aspects of dynamic experiences, language may also lead to different consolidation processes and ultimately, different event representations in memory.

One limitation of the present study is between-subject experiment design. Different participants were tested in the immediate group, wake group and sleep groups, which possibly lead to individual difference performance in memory recall. This is a common problem of between-subject design and in sleep studies in particular, as it is not possible the test the same group in the immediate and sleep conditions without introducing additional experimental confounds. However, the language effects in reproduced duration were replicated. Therefore, the observed language effects are unlikely to be due to the better memory or better linguistic skills in the sleep group. In addition, participants generally recollected details of the animations with 99% accuracy in all groups. Therefore, their memories should not differ significantly.

CHAPTER 5

The role of event structure in event segmentation

In Chapter 3 and Chapter 4, I have addressed the question of how language shapes memory for event duration, in particular, whether language influences are observed at memory encoding, retrieval or consolidation. I have proposed a recall-based account and highlighted the important role of consolidation in integration of episodic memory and schematic knowledge. However, does language indeed contribute nothing at encoding stage? Is it possible that our reproduction measure is not sensitive to the language-induced conceptualization but other measures are? Note that the linguistic concept at encoding not only conveyed temporal event information, but also other event information, e.g., the goal-subgoal relationships or event aim. Moreover, according to Chapter 1, prior knowledge guides event segmentation and memory. Thus, if our linguistic phrases provide a way to conceptualize the movement of geometric figures, it should be expected that the phrases would influence event encoding via event segmentation. Can the event structure conveyed by language labels affect event encoding then? As reviewed in Chapter 1, event segmentation is often cued by goal-subgoal relationship, as knowledge of what the agent intends would imply the steps needed to accomplish the goal. In this exploratory chapter, I examine the influence of language on event representations, and specifically examine how the event structure implied by language influences the identification of event segments. Since segmentation then determines event memory, perhaps the influence of language on memory does not immediately apply to event duration but it may still play a role on other aspects of event memory, such as the event structure built as a result of segmentation.

5.1 Overview

In our everyday lives, we live in a continuous stream of information and we make sense of this dynamic and complex stream of experience by segmenting it into meaningful sub-events (Zacks & Swallow, 2007; Zacks & Tversky, 2001; Zacks, 2004). According to event segmentation theory, when we parse ongoing experiences into discrete events, we automatically construct an event model based on the perceptual changes and our semantic knowledge. This event model enables us to comprehend the current situation and predict future changes. In addition, when the current event model fails to predict changes, we automatically update it.

It has been suggested that activities experienced in our daily lives are by nature goal-directed (Ghosh & Gilboa, 2014; Williamson & Markman, 2006; Woodward, 1998; Loucks & Meltzoff, 2013; Lichtenstein & Brewer, 1980). Besides, the event aim has the power to cue event segmentation and people tend to perceive an event boundary when they perceive the completion of goal or sub-goals (Zacks & Swallow, 2007; Zacks, 2004). Moreover, it was also argued that goals are hierarchically organized and the goal/sub-goal relationships also contribute to cueing hierarchical segmentation (Zacks & Tversky, 2001). To be specific, by identifying the goal/sub-goal relationships, events can be segmented into a range of temporal grains, with coarse segments comprising multiple lower-level fine-grained segments. For example, the event ‘cooking potatoes’, can be decomposed into sub-events such as preparing the potatoes, frying or boiling them and possibly placing the cooked potatoes in a container. In addition, preparing the ingredients can be further divided into 'peeling potatoes', 'slicing potatoes' and so on. The hierarchical segmentation is guided by the goal/sub-goal relationships of the event. The completion of higher-level goals is associated with coarser segments and the completion of lower-level goals is associated with fine segments (Zacks & Swallow, 2007; Zacks & Tversky, 2001; Zacks *et al.*, 2001).

In reading and communication, we use linguistic phrases such as 'building a house' to refer to events. By using these linguistic concepts, the knowledge of event aim which is stored in the semantic memory is usually activated and mapped onto a new situation to guide event segmentation (Hard *et al.*, 2006; Zacks *et al.*, 2001; Zacks, 2004). It has been found that prior conceptualizations of events lead to coarser segmentation (Newtson, 1973; Vallacher & Wegner, 1987; Zacks *et al*., 2001) as the linguistic concepts will activate the goal-hierarchy stored in the semantic memory. Hard *et al*. (2006) found that exposure to linguistic conceptualizations (for example, hide and seek) resulted in fewer segments for the stimulus movies. It was explained that linguistic phrases activate the knowledge of goal-hierarchy so that the fine units can be grouped into larger chunks. In addition, it has been proposed that people tend to parse events in terms of the most superior goals that they can recognize (Wilde, 1978b, 1978a). Conceptualization will therefore lead to coarser segmentation.

However, the conceptualization effects in event segmentation are inconsistent. Zacks (2004) failed to observe coarser segmentation as a function of an agent's goals. In that study, participants were asked to segment abstract animations of geometric figures which were described as moving randomly or as truly intentional (for example, ‘playing a video game’), but he did not replicate the tendency for coarser segmentation in the intentional condition. Even so, it is worth noting that in Zacks’s (2004) study, even though the event knowledge became available through a linguistic concept, the event itself (playing video games) was by nature relatively random and could not be organized hierarchically according to the goal/sub-goal relationships. As a result, the participants could not chunk fine segments into coarse segments according to the goal-hierarchy. Nevertheless, the events used in Hard *et al*. (2006)’s study, such as ‘hide and seek’, can be organized hierarchically and chronologically under the superior level of the aim. Therefore, as the goal-hierarchy becomes available through linguistic concepts, coarser segmentation could be observed in their study (Hard *et al*., 2006).

As discussed above, conceptualization might not lead to an awareness of an event aim or access to goal/sub-goal relationships because a goal-hierarchy is not present in all types of event (Ghosh & Gilboa, 2014). There are, however, situations in which a specific goal or endpoint can hardly be inferred from linguistic concepts (for example, planes flying in the sky, a snail crawling on leaves). In this situation, the verbs emphasize the actions or movements occurring in the events (flying, crawling), rather than suggesting a fixed endpoint or a specific procedure. Moreover, for events not indicating a specific aim, the recognized segments would be the changes in direction or speed, which cannot be organized hierarchically. Although many studies have investigated the influence of event aim on segmentation, little research has distinguished between events indicating a specific aim (a goal-specific event) and events not referring to a specific aim (a goal-unspecific event).

In general, goal-specific events inherently differ from goal-unspecific events in goal-hierarchy, which could possibly explain the inconsistent conceptualization effect on event segmentation found in previous studies. In the current experiment, the aim was to distinguish goal-specific events from goal-unspecific events and examine whether knowing the event aim would affect event segmentation. To answer that question, following the previous experiments, we tested the event segmentation for fourteen abstract animations (from the experiments reported in Chapter 3 and Chapter 4) when the linguistic concepts were provided (linguistic condition) or were not provided (perceptual condition). Of the fourteen animations, seven were categorized as goal-specific events referring to a specific aim (such as taking the ambulance to hospital) and the other seven were goal-unspecific events not indicating a specific aim (such as planes flying in the sky). When a linguistic conceptualization was available, the participants could conceptualize the abstract animations into meaningful events (linguistic condition). In contrast, when linguistic descriptions were not accessible, the participants would interpret the animations as geometric figures moving around, and event representations would be purely based on perceptual changes (perceptual condition). So, in general, two factors were manipulated: event type (goal-specific events as opposed to goal-unspecific events) and the accessibility of schematic knowledge (linguistic condition as opposed to perceptual condition) and each animation had four conditions: goal-specific events in the linguistic condition, goal-unspecific events in the linguistic condition, goal-specific events in the perceptual condition and goal-unspecific events in the perceptual condition.

Note that the conceptualization effects may arise for different reasons. For example, the level of specificity conveyed by the language may prompt people to conceptualize the events differently, e.g., making dinner is a more generic description than frying an egg. If the linguistic description is more detailed, participants attention maybe guided to the fine-grained segments and perceptual details, which may lead to finer segments rather than coarser segments. In the current study, however, following previous study, we will only focus on the coarse linguistic description (linguistic concept) used in previous chapters, which have been shown to fit the animations well. Since we compare the presence vs. the absence of a concept, we expect to observe coarser segmentation when the event's goal-subgoal relationship becomes accessible via linguistic concepts. In contrast, perceptual segmentation is likely to be fine-grained, as determined by visual changes.

Moreover, in order to maximize the opportunity to observe the event aim effects on segmentation, the participants were asked to identify the most meaningful and natural segments they perceived without any specific requirement for fine-grained or coarse segmentation (largest or smallest units). It should be noted that in the previous experiments, because of the strict requirements of the grain of segmentation, the participants had to make an effort to figure out the largest and smallest units in the events. It has been suggested that the completion of goals usually coincides with a burst of physical changes, so physical changes and goal-hierarchy possibly entail a similar hierarchical structure (Hard *et al*., 2006; Zacks, 2004). So, if the participants were explicitly guided to identify the fine or coarse segments, similar segmentation could possibly be achieved, with or without the knowledge of intentions or event structures. However, if the participants were not explicitly guided to segment into fine-grained or coarse-grained segments, the event segmentation would be driven by their own intuition or preference. When the goal-hierarchy can be inferred from the linguistic concepts, coarse event boundaries would be predicted by both the conceptual cues (goal completion) and the physical cues, whilst fine-grained segments would only be predicted by physical changes. So being exposed to the linguistic concepts would make the coarse segments more salient to the participants and they would be more likely to recognize the coarse segments as the most natural and meaningful units.

Following the previous findings, it was predicted that the influence of linguistic concepts would vary as a function of event type. To be more specific, it was expected that coarser segmentation would be observed for goal-specific events (such as making a bed) when the abstract animations are conceptualized into a meaningful goal-specific event (linguistic condition), as the participants could segment according to goal-hierarchy and chunk the fine segments into larger units. However, for the events without a specific aim (such as planes flying in the sky), the goal-hierarchy structure would be unavailable regardless of the presence of linguistic concepts. Presumably, the detection of event segmentation would primarily be directed by the perceptual changes (such as changes in direction or speed) with relatively less influence from linguistic concepts such as the completion of sub-goals. Conceptualizing an abstract animation into a goal-unspecific event would therefore not elicit coarser segmentation.

5.2 Experiment 1

As described in the introduction, the present study was designed to test the natural and meaningful segments in abstract animations which were conceptualized (linguistic condition) or not conceptualized (perceptual condition) into goal-specific or goal-unspecific events. In addition, it was predicted that interaction effects would be observed between event aim and semantic description and that the influences of conceptualization on event segmentation would vary as a function of event aim.

5.2.1 *Method*

Participants

A total of 87 English native-speakers were recruited online. They were given a small cash payment in recognition of their participation. On the 87, 43 participants were randomly allocated to the perceptual group (no linguistic conception), and 44 were assigned to the linguistic group (with linguistic conception). However, four participants in the perceptual group and four in the linguistic group were excluded from the data analysis as their mean numbers of segments were identified as outliers within their group, using a stem-and-leaf plot. In total, therefore, there were 39 participants in the perceptual group and 40 in the linguistic group.

Materials

This study used the same 21 abstract animations from the experiment reported in Chapter 3, except for the manipulation of linguistic concepts. The 21 ambiguous cartoon-like animations varying in duration (from three seconds to nine seconds) were composed of simple geometric figures moving around. Each animation was paired with a single linguistic phrase providing information about who and what the geometric figures were meant to represent (note there was no speed manipulation in this study). In the current study, only one version of the linguistic descriptions was used (half fast version phrases (11 phrases)and half slow version phrases (10 phrases)).

Goal-specificity test

To identify whether the phrases indicated explicit aims, the goal-specificity of all the 21 animations was tested. Using the results of this test, the seven animations with the highest goal-specificity score and the seven with the lowest goal-specificity score were selected, which were considered as goal-specific events and goal-unspecific events respectively. The goal-specificity test questionnaire was constructed as an online questionnaire. In the questionnaire, participants learned about the animations together with the corresponding linguistic descriptions. After learning each animation, they had to judge whether the event described by the linguistic description should be considered as an achievement or accomplishment. They indicated their answers by ticking their chosen option, Yes or No. Two examples were provided in the instruction: the animation of ‘a snail crawling on leaves’ was suggested as ‘not an accomplishment’ and the animation of ‘the plumber fixing a leak’ was suggested as ‘an accomplishment’, as the event indicated the accomplishment of a clear aim. The percentage of Yes judgments to each animation across all the participants was regarded as the degree of goal specificity, which reflected the general pattern of the participants’ judgment of the existence of explicit aims for the events.

Notably, the judgments of goal-specificity of the animations were not consistent across participants. For some animations, the agreement was relatively low (*see* Table 5.1). For example, for animations 4_2, 4_3, 5_2 and 6_1, almost half of the participants considered the animation as an achievement whilst the other half considered it as not an achievement. The goal-specificity scores for the 21 animations were therefore ranked from lowest to highest, and then the 21 animations were divided into three equal groups in terms of the goal-specificity score. The seven animations with relatively low agreement were discarded. The seven animations with the highest goal-specificity were categorized as goal-specific animations and the seven with the lowest goal-specificity score were categorized as goal-unspecific animations. We checked that the number of fine-grained segments collected in Chapter 3 for our stimuli as well as their clock duration. The average number of fine segments for goal-specific events and goal-unspecific events were 4.93 vs. 2.77 respectively. The average number of stimuli for goal-specific events and goal-unspecific events were 5.86s vs. 6.26s.A comparison of the clock duration across the two groups was not significantly different (p > .10).

Procedure and design

Event segmentation for all the 21 ambiguous animations from the previous experiment were tested in the conditions that linguistic descriptions were provided (the linguistic group) or absent (the perceptual group). Of the 21 animations, there were fourteen experimental animations (seven goal-specific events and seven goal-unspecific events) and seven fillers. To be specific, the participants in the linguistic group were exposed to all 21 ambiguous animations together with the corresponding linguistic concept – a phrase describing the event occurring in the animation (for example, a Chinese lantern rising up into the sky), which related the abstract animations to real-life experiences. Hence, with the linguistic description, the participants would interpret the animations as meaningful events. In contrast, the participants in the perceptual group watched the ambiguous animations without exposure to linguistic phrases, so that they would consider the animations as geometric figures moving around randomly or would form their own understanding of the animation.

In the segmentation questionnaire, the same instructions and examples were provided for both the linguistic and the perceptual groups, and the participants were asked for the most natural and meaningful units they could identify. In other words, they were instructed to calculate how many natural and meaningful units appeared in the animations. They were guided to identify the event boundaries in the animations (the points at which one event ends and another starts) and count how many times event boundaries occurred in each animation. They were asked to type their answers in blank boxes. They were allowed to watch the videos as many times as they wanted to.

Table 5.1 Experiment 1: *Response to the aim-directness question*

Duration	Categorization	Phrase	fine segment	Percentage of participants	
				Achievement	Not achievement
3s	Goal-unspecific	Planes flying in the sky	3.3	16.13%	83.87%
7s	Goal-unspecific	Visitors rowing in the river	2.6	19.35%	80.65%
9s	Goal-unspecific	A tourist driving a car on a mountain road	2.4	19.35%	80.65%
3s	Goal-unspecific	A Chinese lantern rising into the sky	2.2	22.58%	77.42%
7s	Goal-unspecific	A person stepping onto a frozen lake	2.9	22.58%	77.42%
9s	Goal-unspecific	Goldfish swimming in an aquarium tank	3.8	22.58%	77.42%
3s	Goal-unspecific	An army tank going downhill	2.2	25.81%	74.19%
5s	Removed	Two friends jogging to meet in the park	4.4	32.26%	67.74%
4s	Removed	Someone drinking a glass of beer	3.2	48.39%	51.61%
4s	Removed	Products in a factory’s assembly line	3.5	48.39%	51.61%
6s	Removed	Mixing ingredients in a bowl	5.0	51.61%	48.39%
8s	Removed	Horses sprinting at a race course	7.1	58.06%	41.94%
6s	Removed	Athletes competing in a hurdle race	5.7	61.29%	38.71%
8s	Removed	Students running up to their classroom	5.7	64.52%	35.48%
4s	Goal-specific	Cutting a banana with a knife	4.1	67.74%	32.26%
8s	Goal-specific	Taking the school bus home	4;1	67.74%	32.26%
7s	Goal-specific	A child running downstairs to find mom	6.6	74.19%	25.81%
9s	Goal-specific	Grandma taking the bus to the hospital	6.2	77.42%	22.58%
5s	Goal-specific	A man swimming across a river	5.9	80.65%	19.35%
6s	Goal-specific	Students in a relay competition	4.0	87.10%	12.90%
5s	Goal-specific	An old man climbing a tree to pick a fruit	3.6	90.32%	9.68%

Data Treatment and statistical analyses

One dependent measure was used for the analysis: the number of natural segments. To minimize the influence of outliers in the analysis, the segmentation beyond three standard deviations from the item mean was removed (a total over all subjects of about 3% of the data points).

The hypothesis was tested using linear mixed-effect models in the lme4 package of R (R core team, 2017) (Bates *et al.*, 2015). The models included the maximal random effects structure when the model can converge (the random intercepts for subjects and items, and by-subject random slopes for event aim) (Barr *et al.*, 2013). Effects of interests (independent factors), goal specificity (goal-unspecific event as opposed to goal-specific event) and linguistic concepts (perceptual condition as opposed to linguistic condition) were included in the final model as fixed factors. To test the effects of interest, the full model with the effect of interest was compared with a model without this effect.

5.2.2 Results

Table 5.2 Experiment 1: Summary of the fixed factors from the linear mixed-effects model for predicting the number of segments for the animations.

Fixed effects	*Estimated coefficient*	*Standard error*	*t-value*	*p-value*
Intercept	2.12	0.21	10.33	<.001
Event aim (Goal-specific / goal-unspecific)	2.25	0.43	5.22	<.001
Event aim X semantic description	-0.80	0.37	-2.17	0.037

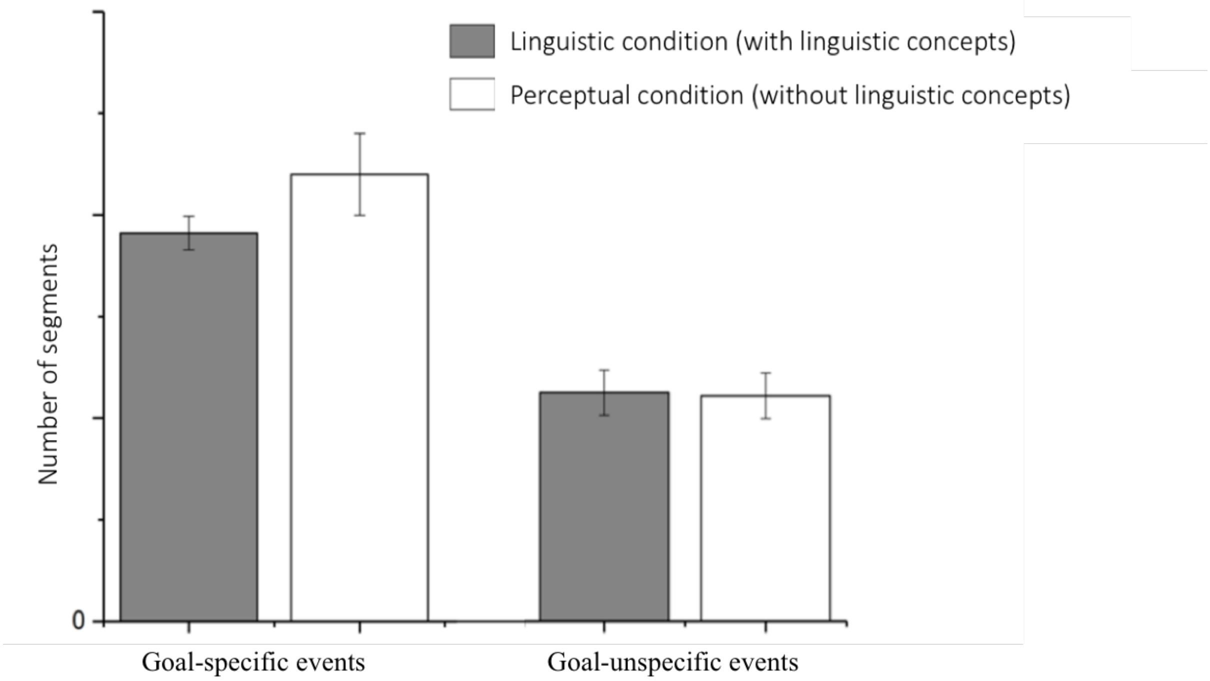


Figure 5.1 Experiment 1: The number of segments varying as a function of event aim and linguistic concepts. Error bars indicate standard errors for each group.

Mixed-effect modeling revealed a main effect of event aim ($\chi2(1) = 18.31$, $p<0.001$). It was shown that people tended to report more segments when the events indicated explicit aims (goal-specific events) than when events did not refer to an aim (goal-unspecific events). However, there was no significant main effect of linguistic description ($\chi2(1) = 0.3$, $p = 0.58$). Therefore, the main effects of linguistic descriptions were removed from the full model. The absence of a linguistic effect suggests that in general the description did not alter the way in which participants segmented the events.

Importantly, a significant interaction was found between linguistic descriptions and event aim specificity ($\chi2(1) = 4.47$, $p = 0.034$) suggesting that description influenced event segmentation differently as a function of event aim (*see* Figure 5.1). The model is summarized in Table 5.2. To further explore the interaction, planned comparisons were performed and the results showed that the number of segments for goal-specific events was significantly decreased in the linguistic condition relative to the perceptual condition (t(77) = 2.67, p = 0.009). Nevertheless, for the goal-unspecific events, there was no significant difference between the linguistic condition and the perceptual condition (t(77) = 0.89, p = 0.38).

5.2.3 Discussion

The results were consistent with the predictions. It was found that the effects of conceptualization in event segmentation were modulated by event aim. In other words, conceptualizing an abstract animation into a goal-specific event did elicit coarser segmentation, whereas event segmentation did not vary significantly if the abstract animations were conceptualized into goal-unspecific events. The results have implications for the inconsistent conceptualization effects observed in the previous studies (Newtson, 1973; Zacks, 2004; Hard *et al*., 2006). When goal-hierarchy was accessible, finer segments could be chunked into larger units according to goal/sub-goal relationships, which led to coarser segmentation for goal-specific events in the linguistic condition relative to the perceptual condition. Nevertheless, when the goal/sub-goal relationships were absent inherently (goal-unspecific events), being exposed to linguistic concepts made no difference to event segmentation.

It was also observed that there were significantly more segments for goal-specific events than for goal-unspecific events. On the one hand, this was possibly because of the selection of the stimuli (for example, the goal-specific events were generally longer). It is also possible that goal-specific events by nature contain more changes than goal-unspecific events. The explanations and implications of these findings will be discussed later in this chapter in the general discussion.

It has been suggested that during event segmentation, by making use of schematic knowledge, we automatically construct an event model to represent the current situation, which enables us to make predictions for future changes (Zacks & Swallow, 2007; Zacks & Tversky, 2001). The present findings have shown that linguistic concepts affect event segmentation differently for goal-specific events and for goal-unspecific events. It is therefore possible that linguistic concepts contribute differently to making predictions for goal-specific events and goal-unspecific events during event segmentation. For goal-specific events, linguistic descriptions would indicate the possible steps in the events and their temporal sequences by making use of prior knowledge. For goal-unspecific events, however, event schemas do not allow us to anticipate possible changes, so the detection of event segments must rely primarily on the bottom-up sensory characteristic. To test this idea, the predictability of goal-specific events and goal-unspecific events was examined in Experiment 2.

5.3 Experiment 2

To examine whether conceptualization contributes more to making predictions for goal-specific events than for goal-unspecific events, a questionnaire survey was employed. In the questionnaire, participants were guided to rate the extent to which a linguistic description or phrase helped them to predict changes in the animations. It was expected that linguistic concepts would play a larger role in cueing changes for goal-specific events than for goal-unspecific events.

5.3.1 Method

Participants

Thirty-two English native speakers from ‘Prolific’ participated in the experiment in exchange for a small payment.

Materials

The experimental stimuli consisted of 21 abstract animations and corresponding linguistic descriptions, which was the same as in Experiment 1.

Procedure and design

The predictability of linguistic concepts for all 21 ambiguous animations used in Experiment 1 was tested. As in Experiment 1, there were fourteen experimental animations (seven goal-specific events and seven goal-unspecific events) and seven fillers. In addition, in this test, the abstract animations were presented together with the corresponding linguistic phrases and the participants could watch the videos as many times as they wished.

For the predictability test, participants were asked to judge the extent to which the linguistic description helped them to predict changes in the animation on a scale from 1 to 7, with 1 indicating ‘not helpful at all’ and 7 indicated ‘very helpful’. Two examples were provided in the instructions: the animation of ‘a snail crawling on leaves’ was suggested to rate as not very helpful (1 or 2), as the path of movement was relatively random and could not be inferred from the linguistic description; and the animation ‘the plumber fixing the leak’ was suggested to be rated as ‘very helpful’ (6 or 7) as the linguistic description enabled the sub-events in the animations to be predicted, such as the plumber would go to the leak position first and then fix it.

5.3.2 Results

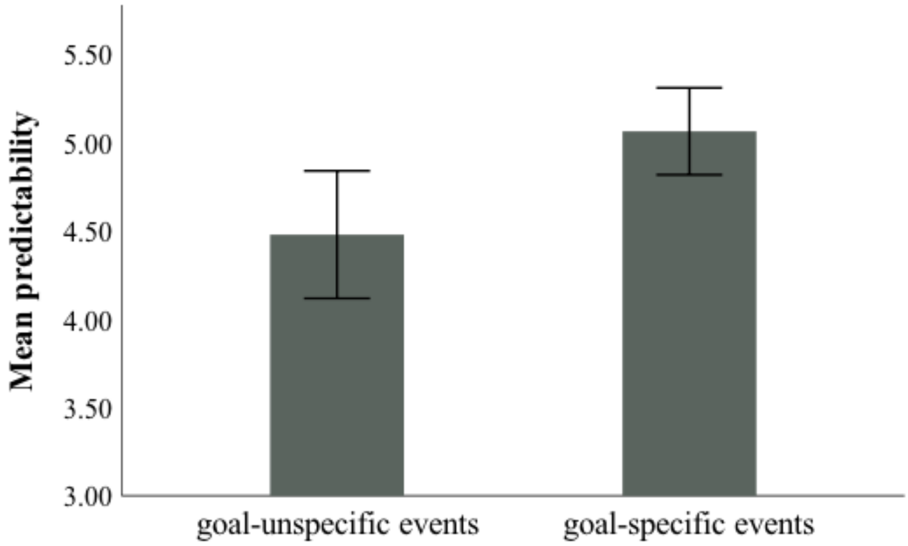


Figure 5.2 Experiment 2: The predictability score for goal-unspecific events and goal-specific events. Error bars indicate standard errors for each group.

A paired-sample T-test revealed that linguistic concepts were rated as more helpful for predicting sensory changes for the goal-specific events than for the goal-unspecific events (t (31)=-5.01, p<0.001). The results showed that schematic knowledge played a larger role in cueing segmentation for goal-specific events than for goal-unspecific events.

5.3.3 Discussion

Consistent with the predictions and with previous findings, linguistic concepts were rated as more helpful for predicting changes for goal-specific events than for goal-unspecific events. This showed that schematic knowledge of goal-specific events suggested possible procedures or temporal sequences which made it possible to make predictions of future changes. Event schemas of goal-unspecific events, however, did not enable the participants to infer possible procedures and steps, so the detection of event boundaries had primarily to rely on the bottom-up perceptual changes.

5.4 General discussion

The present study was designed to explore whether goal-specific events differ from goal-unspecific events in natural event segmentation as a function of the presence or absence of linguistic conceptualization. Consistent with the predictions and with previous findings, it was found that exposure to linguistic descriptions led to coarser segments, but only when the animations were conceptualized into events indicating explicit aims (the goal-specific events in the linguistic condition). Nevertheless, the presence or absence of a linguistic conceptualization did not modulate event segmentation for goal-unspecific events. The results show that linguistic conceptualization influences event segmentation differently as a function of an event's explicit aim. In line with this finding, in Experiment 2, linguistic concepts were rated as more helpful for predicting changes for goal-specific events than for goal-unspecific events. Interestingly, more segments were observed in the goal-specific events than the goal-unspecific events, regardless of whether linguistic descriptions were provided.

The fact that conceptualization effects were only observed for goal-specific events indicated that schematic knowledge had a larger influence on event segmentation for goal-specific events than for goal-unspecific events. This result is consistent with the previous finding that events were segmented in a coarser manner when the observers were aware of the event aim (Newtson, 1973;

Vallacher & Wegner, 1987; Hard *et al*., 2001). The different conceptualization effects for goal-specific events and goal-unspecific events could stem from the difference in internal event structures. To be specific, goal-specific events (such as making a bed) can be organized according to goal-hierarchy (Kurby & Zacks, 2008), whilst goal-unspecific events by their nature lack a goal-hierarchical structure. When no linguistic descriptions were provided, the event schema could not be related to the animations' geometric figures or was partially inferred from the scene setting (trapezoid shapes on a large blue shape might suggest boats on water). In either case, event segmentation appeared to have been driven by bottom-up perceptual changes in the animation with relatively little influence from schematic knowledge. However, when linguistic descriptions were provided, the geometric figures would be conceptualized as meaningful real-life events and schematic knowledge would be activated. In this condition, the knowledge of goal-hierarchy from an event schema can be mapped onto the new situation which will enable the fine-grained segments to be chunked into larger units accordingly. Also, as Wilde (1978b, 1978a) suggested, people tend to segment according to the most superior level of goal that they can recognize. Therefore, people tend to apply coarser segmentation when the events are conceptualized into goal-specific events. For goal-unspecific events, however, because of the absence of an inherent goal/sub-goal hierarchy in schematic knowledge, fine-grained segments could not be clustered into coarser units according to the goal/sub-goal relationships. Hence, mapping meaning onto the events would not lead to coarser segmentation as the superior level of aim is still unavailable.

Even though conceptualizing abstract events into goal-specific events did indeed lead to fewer segments in the experiment, this does not necessarily imply that the number of segments being perceived is decreased by conceptualization. It is possible that the importance or salience of the identified event boundaries is changed according to a goal-hierarchy. Since conceptual breaks (such as the completion of sub-goals and goals) are usually compounded with a burst of sensory or perceptual changes, the same event boundaries or segments could be recognized on the basis of the sensory characteristics of the events, regardless of access to schematic knowledge (Hard *et al*., 2006). However, the importance or salience of the identified event boundaries could be altered as a function of event conceptualization. Thus, in the condition in which no linguistic concepts were provided, the salience or importance of event boundaries were judged according to the quantity or quality of the perceptual changes (changing colour, position, speed and so on). The larger or more obvious perceptual changes would be recognized as the coarser segments. In contrast, in the condition in which phrases were provided, the goal hierarchy was mapped onto the events. In this explanation, the salience or importance of event boundaries is cued by both the goal hierarchy and the perceptual changes. The event boundaries relating to the completion of the superior-level goal would be considered as more important or more salient than the event boundaries relating to subordinate goals, whilst the perceptual changes not relating to goal completions (which could not be predicted from goal/sub-goal relationships) would be recognized as the least important. This account is consistent with previous findings suggesting that coarse event boundaries are usually associated with conceptual knowledge, whereas fine segments are usually predicted by sensory changes (Zack, 2004; Hard *et al*., 2006). Conceptualizing ambiguous events into goal-specific events would therefore increase the salience of the coarse segments, which would make the importance of event boundaries vary with goal-hierarchy. In the current experiment, participants were asked to identify the most natural and meaningful segments in the animation without an explicit requirement about the size of the units. Hence, participants' answers would be influenced by their judgment of the importance or salience of the perceived segments rather than the grain of segmentation (Zack, 2004; Hard *et al*., 2006). In this view, as the conceptualization of the goal-specific events would make the coarse event boundaries more salient or more important, fewer segments would be reported for goal-specific events due to conceptualization.

These findings can possibly explain the inconsistent previous findings of conceptualization effects in event segmentation. It is suggested that because Zack (2004) used goal-unspecific events which lack inherent goal-hierarchy, he failed to capture the influence of conceptualization in the degree of hierarchical segmentation. However, it is also possible that segmentation according to unit size diminishes the power to detect conceptualization effects in event segmentation, as conceptualization does not necessarily affect the number of identified event boundaries but alters the importance or salience of the identified segments. Hard *et al*. (2006), however, also asked participants to segment into fine or coarse grain, but they did observe conceptualization effects. The inconsistent findings are therefore more likely to be due to the differences in events' hierarchical structure.

The finding from Experiment 2 showed that linguistic concepts were rated as more helpful for predicting sensory changes for goal-specific events than for goal-unspecific events. According to event segmentation theory, the predictability of future units determines segmentation. When predictions of upcoming material based on prior knowledge are not possible, a boundary is perceived. Because goal-specific events are intended to accomplish an explicit goal, the possible steps in the events and their temporal sequences can easily be inferred by making use of prior knowledge. Linguistic descriptions therefore largely influence predictions for goal-specific events, but for goal-unspecific events, event schemas do not allow possible perceptual changes to be anticipated, so the detection of event segments is primarily driven by the bottom-up sensory characteristic. As a result, schematic knowledge played a larger role in cueing segmentation for goal-explicit events than for goal-unspecific events and the access to linguistic descriptions made little difference to event segmentation.

Another important finding is that significantly more segments were observed in goal-specific events than in goal-unspecific events. One possible reason for this is that the animations' clock durations were significantly longer for goal-specific events than for goal-unspecific events. However, an independent T-test ($p>.05$) across stimulus durations in each event type condition indicated no significant difference, so the difference in the number of segments was not due to the animations' actual durations. Another possibility is that goal-unspecific events and goal-specific events might inherently contain different numbers of perceptual changes. Jeunehomme and D'Argembeau (2018) found that spatial displacements (such as walking to the library) were associated with fewer segments than goal-directed events (such as posting a letter). In the stimuli used in the current study, all of the goal-unspecific animations and five of the goal-specific animations referred to spatial displacement, so the goal-unspecific events included more spatial displacement than the goal-specific events, which possibly led to fewer perceptual changes occurring in the goal-unspecific events. Further research is required to test this possibility.

One limitation of the current study is that the numbers of segments for goal-specific events and for goal-unspecific events were only explored in an off-line segmentation task, rather than on-line. The participants first observed the animations and then counted the number of units, whereas in other studies the participants had to press a key as the animation unfolded. So a stronger test of the current findings would be to test on-line event segmentation. This would enable the positions at which viewers perceive boundaries and their hierarchical alignment in time to be examined (Zacks, 2004). Also, the numbers of changes in the animations were not manipulated in the current study. Ideally, a study could be designed in which the same abstract videos are viewed as goal-unspecific events and as goal-specific events so that the influence of sensory changes could be better controlled. Finally, it is widely accepted that event segmentation affects subsequent memory (Sargent *et al.*, 2013; Swallow *et al.*, 2009; Zacks *et al.*, 2007). It would be interesting to test how the accessibility of event aim influences memory of an event, thus shedding some light on the underlying mechanisms of event processing.

In summary, in the current experiment, it was found that conceptualizing ambiguous events into meaningful events elicited coarser segmentation only when the linguistic concepts indicated an explicit aim or endpoint. This finding shows that schematic knowledge contributes more to event segmentation for goal-specific events than for goal-unspecific events. The results also suggest the possibility that aimless events tend to be associated with fewer perceptual changes relative to goal-specific events. Further studies are necessary to explore the influence of goal-hierarchy on on-line segmentation and how it contributes to memorizing an event.

CHAPTER 6

CONCLUSION

6.1 Summary of the findings

We live in a constant stream of time and tend to keep track of event duration in language comprehension and real-world experience. In this thesis, I have investigated how we represent event duration in memory and language, and how linguistic concepts or schemas modulate duration representation. In Chapter 2, I examined how we represent event duration in language. Sentences describing the same events were modified by temporal adverbials indicating longer or shorter durations either on a small-time scale (for example, playing basketball for ten minutes as opposed to playing basketball for an hour) or on a large-timescale (travelling for a week as opposed to a year). Consistent with previous findings, it was found that longer events took longer to process. In addition, I explored the associated background knowledge of events varying in duration and found that consistent with previous findings, longer events were associated with relatively more diverse information. The results indicate the possibility that duration representation in language is rooted in the type of background knowledge activated during processing.

In Chapters 3 and 4, I investigated how people retrieve event duration from episodic memory and how the reproduced durations are influenced by event schemas. I tested duration reproduction and verbal recall for abstract animations which were conceptualized into fast or slow events through linguistic phrases indicating fast or slow speed (for example, running downstairs as opposed to crawling downstairs). It was found that the amount of verbally recalled information predicted the reproduced duration; the more recollected information there was for an event, the longer the reproduced duration. It was proposed that recalled information density (the amount of recalled information per time unit) explained duration reproduction. Higher recalled information density led to over-estimation of duration reproduction whereas lower recall information density resulted in under-estimation of reproduced duration. With respect to the role of schemas in duration memory, I investigated whether and when linguistic concepts shape the memory of event durations. It was found that, if memory recall happened shortly after learning, linguistic concepts biased reproduced duration only when the duration reproductions were cued by the linguistic concepts. However, if memory recall was taken after a twelve-hour delay with sleep, linguistic concepts biased duration reproduction even though the linguistic concepts did not mediate the recall process. These results suggest that episodic details and linguistic concepts are stored independently at initial encoding and can be combined during recollection or gradually merge into each other during sleep.

In Chapter 5, I explored how prior knowledge of schemas and event structures modulates event segmentation, the building blocks of event memory. I tested whether conceptualizing abstract animations through language into goal-specific events or goal-unspecific events would influence the grain of segmentation. In the experiment, I distinguished goal-specific events (events with a specific aim, such as making a bed) from goal-unspecific events (events without a specific aim, such as a snail crawling on leaves). It was found that conceptualization led to coarser segmentation only for goal-specific events. It was explained that the goal-specific events could be organized according to a goal-hierarchy and that fine segments could be chunked into larger units (Kurby & Zacks, 2008; Zacks, 2004; Hard *et al.*, 2006), whereas goal-unspecific events by their nature lack a goal-hierarchical structure, so being exposed to linguistic concepts would not elicit coarser segmentation. It was suggested that schematic knowledge contributes more to event segmentation for goal-explicit events than for goal-unexplicit events because of the differences in events' hierarchical structure.

I explored duration representations in language and memory and elucidated some of the underlying mechanisms in duration representation. I suggest that duration representations are grounded in the knowledge activated during processing, which could result from the combination of episodic details and semantic information. In addition, I discussed the role of schematic knowledge in processing novel situations: schematic knowledge mediates language comprehension, event segmentation and memory recall and interacts with episodic memory at recall and during sleep. The findings contribute to event representation, duration representation, schema theory and event segmentation theory. This will be discussed more fully and an information-activation based model will be suggested later in this chapter.

6.2 Representing event duration in language

This study is thought to be the first to investigate representations of explicit temporal information in language comprehension. In Chapter 1, I reviewed previous studies of the representation of event duration in language. These studies have mainly focused on the influence of temporal shifts and their role in cueing mental model updating, but there had previously been only a few studies which considered the representations or comprehension of event duration in language, and they had manipulated event duration using different verbs (hunting a deer as opposed to shooting a deer) (Coll-Florit & Gennari, 2011). However, different verbs would possibly activate different event schemas (for example, hunting and shooting), so even though it was observed that longer events took longer to process, the different processing costs associated were probably driven by different event schemas, rather than by event duration itself. In this study, this compounding effect was better controlled by manipulating temporal adverbials for the same events on different time-scales.

Previous theories had focused on emphasizing the fact the mental models contain temporal information, as in the indexing model (Zwann & Radvansky, 1998; Zwaan *et al.*, 1998). Other accounts have argued for some version of mental simulation whereby the events described by the language are mentally simulated to unfold in the mind (Lindsay & Kamide, 2013; Barsalou, 1999; Zwaan, 2004). In contrast to these approaches, I suggest that duration representation in language stems from the variability and diversity of the associated knowledge. Specifically, in language comprehension, temporal adverbials serve as a cue to activate possible sub-events based on prior experience and schematic knowledge. I argue that more diverse information was activated for longer events than for short events, which led to greater processing costs for longer events.

Mental models and simulation theories do not specify how people simulate an event in a temporally compressed manner. According to the simulation account, people tend to mentally simulate described events so that longer events should be simulated for longer. However, it is obvious that people simulate event duration in a compressed manner. Processing the phrase ‘reading for five hours’ does not need five hours. Simulation theory does not have an explanation for this temporal compression in language comprehension. In contrast, the information activation-based account explains this from the information actually activated by the reader, which can be measured by using association and recall tests.

In general, the story-reading experiments described in Chapter 2 provide insights into representations of explicit temporal information that longer events took longer to process. More importantly, I proposed a novel explanation for the processing cost associated with longer events; the semantic diversity of activated information can explain the observed duration effects in retrieval cost. The present findings thus not only provide novel findings that constrain existing theories but also provide an alternative explanation for the processing cost associated with temporal adverbials. Unlike updating mechanisms proposed by mental model accounts, or more complex mental simulations, it is argued that the nature of the world knowledge being activated during processing is responsible for processing cost.

6.3 Representing event duration from episodic memory

In Chapters 3 and 4, I explored 1) how people reproduce event duration from episodic memory, 2) how language shapes memory of event duration, and 3) the role of sleep in consolidating event memory. Based on these findings, I suggested a recall-based model for duration reproduction, which is different from the traditional model for remembered durations. I shall next discuss the theoretical contributions of these findings to several fields of inquiry.

6.3.1 Contributions to cognitive models of remembered duration

In Chapter 1, I reviewed several cognitive models for remembered durations. All of the models ignored the process of memory retrieval and focused on the encoded stimulus (such as the number of segments identified in stimuli, the complexity of stimuli or the similarity of stimuli). Even though the previous models suggested that people reconstruct duration from memory, they did not propose an objective measure to test memory content directly. Based on the findings of the current study, I suggest a recall-based model and pinpoint the role of memory recall in the reproduction of event duration. Specifically, I suggest that reproduced durations can be explained by the amount of recalled information. This explanation is accompanied by a new measure – the amount of recollected information, in other words, the number of words in the cued recall task.

The recall-based model was shown to be robust and reliable across the experiments carried out in this study and it also has implications for the findings of previous studies. For example, as more complex or more diverse stimuli usually contain a larger amount of information, more information should be recollected during duration reproduction, which lengthens the reproduced duration. This was experimentally manipulated by varying the amount of learning that participants undertook before the time-reproduction task. Because this deeper learning led to the recall of more details in the stimuli, the reproduced duration accordingly lengthened when one viewing was compared with three stimulus viewings. This result clearly demonstrates that the amount of information recalled plays an important role in memory of duration.

The present proposal is similar in spirit to that proposed by Ornstein (1969) in that the complexity of the memory is related to the estimated duration. Other accounts such as the contextual change theory have proposed that the number of contextual changes in the stimuli should explain remember duration (see review in Chapter 1). These accounts capture the finding that more segments in the stimuli leads to longer remembered duration. However, the number of segments is not the only factor contributing to remembered duration, as the details recalled also account for variance over and above the influence of the number of segments. Other studies have also suggested that familiarity of the stimuli and the similarity between events, which encourages hierarchical structuring, also influence remembered duration. Therefore, the present account goes further in providing a single underlying explanation for all these effects. Because familiarity (prior knowledge) ultimately determine how the event is segmented and recalled, the information recalled is the best predictor of remembered duration. Our findings therefore offer a new theoretical approach to explain remembered duration.

The recall-based model also provides possible alternative explanations for widely observed phenomena thus opening up further research avenues. For example, older people are known to experience time passing faster than younger people (Friedman & Janssen, 2010; Janssen *et al.*, 2013; John & Lang, 2015; Block *et al.*, 1998). This phenomenon has been explained as resulting from the novelty of experiences, familiarity or more or less attention being allocated to the stimuli (Friedman & Janssen, 2010; Block *et al.*, 1998). Whereas, according to the theory which I have proposed, it is possibly caused by the fact that as they get older, adults get worse at event segmentation (Kurby & Zacks, 2011) and at remembering episodic details (Levine *et al.*, 2002；Grady *et al*., 2006), so there will be less recalled information about past experiences. As a result, shorter remembered durations and the sensation of time speeding up can be observed. The proposed theory also provides a possible explanation for the illusion of temporal compression. When we think about a past period, especially longer durations (such as a year or ten years), we always get the feeling that time moved faster than it is supposed to. Why do we get the sense that time in memory is dramatically compressed? According to the recall-based account, it is because very limited information can be successfully recalled after one or three years. This forgetting of episodic details leads to great temporal compression in memory recall. It would be interesting to carry out further experiments to test these ideas and explore how memory for event duration varies as a function of aging.

6.3.2 Contributions to temporal distortion in duration reproduction

Following the recall-based account, I propose an explanation for the so-far unresolved Vierordt's law in reproduced duration. This law refers to the tendency to over-estimate short events and under-estimate long ones (Gu & Meck, 2011; Lejeune & Wearden, 2009). I argue that the higher density of recalled information in short events would lengthen duration reproduction whilst the lower density of recalled information in longer events would shorten duration reproduction. The density of the information recalled ultimately stems from segmentation mechanisms so that events with longer segments (as would typically be the case in longer events) tend to be recalled in less detail because within-segment information is not encoded and recalled as well as information at event boundaries. To my knowledge, this is the first study to explain Vierordt's law from the perspective of memory recall.

It should be noted that even though recalled information density predicts temporal compression, we are still unable to disprove the ‘central tendency’ in remembered duration. Central tendency explanations of temporal compression argue that participants simply tend to average the duration of the stimuli which they have observed and that all reproductions therefore tend to be clustered around the mean stimulus duration. Because of this bias to the mean, short stimuli are over-reproduced and long ones are under-reproduced. This explanation is particularly proposed by theories of prospective timing discussed in Chapter 1, such as the scalar model. A stronger test of this explanation account would be to examine whether duration reproduction of the same event (for example, a six-second event) would deviate as a function of the range of stimulus durations observed in the experimental context (one to seven seconds as opposed to five to eleven seconds). Vierordt’s law predicts a central tendency of duration reproduction so that the same event would be over-reproduced or under-reproduced depending on the range of stimulus durations in the experimental context. To be more specific, a six-second event would be over-reproduced if the range of stimulus durations was from five to eleven seconds but under-reproduced if the range of stimulus durations was from one to seven seconds. In contrast to Vierordt’s law, according to our theory, it is expected that the reproduced duration would not change with the range of stimulus durations, as the recalled information for events will possibly remain similar regardless of the range of stimulus durations, unless memory encoding and segmentation change with the experimental context, which is also possible. It would also be worth carrying out further studies to test these theories and reveal the underline mechanisms of temporal distortion.

6.3.3 Contributions to the relationships between language and episodic memory

Previous research has suggested that semantic memory interacted with episodic memory and resulted in schema-mediated distortion in episodic memory, even soon after encoding. However, when and the interaction occurred remained unclear, at encoding or at retrieval. In Chapters 3 and 4, I addressed this issue by testing how linguistic labels appearing at encoding and retrieval affect memory of abstract animations.

In Chapter 3, it was found that a linguistic concept biased duration memory only when the retrieval was mediated by language, if memory of animations was tested immediately after learning. This result suggested the possibility that at the initial encoding stage, episodic memory and linguistic information were stored separately. Then, at the retrieval stage, if the retrieval process was mediated by language, remembered episodic details would be recollected/reconstructed according to the given linguistic information. Hence, schematic knowledge and episodic memory were merged during memory reconstruction. However, when visual frames served as the retrieval cue, memory reconstruction can be achieved by simulating the movement of the visually presented objects without the activation of linguistic concepts. This is why no language effects were observed immediately after learning, unless the linguistic concepts mediate memory retrieval.

These findings were the first to investigate the modulation of linguistic concepts in event memory, and duration memory specifically. As such, they highlight potential differences with object memory, which has mostly been studied so far, and introduce novel facts to be explained in any theory of event memory and language. At least when it comes to duration memory and event reproductions, language influences are not immediate, suggesting that language and visual episodic details are initially kept separate. These results therefore go against interactive encoding accounts arguing that labels may immediately influence the encoded memory. They also support recall-based accounts in that language-cues do indeed influence the retrieved memories. Some of the discrepancies with object studies such as that of Lupyan (2008) might be due to the different tasks used (mental replay vs. recognition memory). But they can also be due to the inherent differences between objects and events. Events are complex entities that change over time and thus our mind encode multiple visual cues that can help retrieval, and thus, do not necessitate language to aid this process. In contrast, static pictures of objects in a context in which many similar objects have observed may be harder to retrieve, and thus, labels may be used to facilitate this process. Whichever the case, our results impose important constraints on existing theories and pave the way to investigate other aspects of event memory that may or may not change as a function of language. Clearly, duration memory was not immediately influenced by linguistic labels.

Is episodic encoding completely independent of linguistic concepts initially? In other words, does the merging of episodic details and linguistic concepts not occur during encoding? The segmentation experiments described in Chapter 5 showed that schematic knowledge affects the number of identified segments. Specifically, it was found that exposure to linguistic concepts elicited coarser segmentation (fewer identified segments) for abstract animations only when the animations were conceptualized into goal-explicit events (such as making a bed). The results suggested the possibility that schematic knowledge possibly influences encoded segmentation. It has therefore been shown that event structures indicated by linguistic concepts were integrated with episodic details at the encoding stage. So why does schematic knowledge introduced at encoding influence event segmentation but not the reconstruction of temporal information?

It should be noted that the number of identified segments in the event segmentation task presented in Chapter 5 were actually the results of judgments of the natural and meaningful segments observed, rather than the number of perceived segments in the absence of instructions. Specifically, as the natural and meaningful segments in events were asked for, the participants had to make a judgment according to the information available. Hence, when linguistic concepts were available, they made judgments based on the goal/sub-goal relationships indicated by linguistic concepts, whereas when the linguistic concepts were absent, they had to rely on perceptual changes in the animations. Therefore, as discussed in Chapter 5, schematic knowledge does not necessarily alter the perceived fine-grained segments or encoded segments, but changes the judgment about what should be considered as meaningful and natural segments.

On the one hand, if schematic knowledge indeed changes the perception of event segmentation, the inconsistent schema effects in event segmentation and duration representations which were reported and discussed in Chapters 3 and 5 were possibly due to the ambiguous nature of the animations. It should be noted that the event structures in the abstract animations were consistent with the goal/sub-goal relationships indicated by the linguistic concepts. Nevertheless, the speed information in the abstract animations was somewhat ambiguous, so the same animations could fit into both fast and slow linguistic descriptions. As a result, the participants could easily map the goal/sub-goal relationships indicated by the linguistic concepts onto the abstract animations at the encoding stage. In contrast, as speed information/temporal information was ambiguous in the abstract animations, it would be harder to map the temporal information indicated by the linguistic concepts onto the time courses of the abstract animations. It is therefore necessary to test whether ambiguity would mediate the integration of episodic memory and semantic memory at the encoding stage.

Another possibility is that the changes in segmentation which were described in Chapter 5, such as the tendency to more coarse segmentation for goal-directed events, were not large enough to cause differences in duration reproductions. Coarser segmentation can still result in similar duration reproductions depending on how much within-segment information is recalled. Moreover, the stimulus animations described in Chapter 3 contained both goal-directed and goal-unspecific events so, on average, there was not enough power to detect an influence of segmentation strategies, if there were any in the experiments described in Chapter 3, to observe a language effect. Thus the relationship between explicit segmentation strategies and duration reproductions remains to be investigated further.

Even though the effects of schemas do not bias recollection of event duration immediately, it was found that schemas introduced at the encoding stage would bias memory for event duration after sleep-mediated memory consolidation. In Chapter 4, it was shown that when the linguistic concepts were only shown at the encoding stage, the schema-mediated distortion of remembered duration could only be observed after a twelve-hour delay with sleep, but not immediately or after a twelve-hour delay without sleep. I therefore suggested that sleep contributed to the integration of schematic knowledge conveyed by the language and episodic memory. Initially, episodic details and linguistic concepts were stored separately in the memory, but after sleep, episodic details stored in the memory were biased towards the features of the event schemas and became schematized or conceptualized. So even though linguistic concepts were not shown at the recall stage, language effects could be observed in memory reproduction after sleep. There is an alternative explanation for the observed schema-mediated distortion in duration reproduction. It has been suggested that episodic memory decays faster than semantic memory (McRae & Jones, 2018). It is therefore possible that during sleep, episodic details are deleted as a consequence of memory decay but that the linguistic information remains in the memory. As a result, when duration is reproduced from memory, linguistic concepts will be used to fill the gaps in event memory. From this point of view, the merging of episodic memory and semantic memory does not take place before the retrieval stage. Instead, the interaction between event memory and schematic knowledge occurs only when the concept helps retrieval. Nevertheless, it is still revealing that the phrases seen at encoding were not forgotten and that details were not lost, and that sleep is necessary to observe language effects, compared with twelve hours of waking time, during which memory is also known to decay. It would be interesting to carry out further experiments to better understand the underlying mechanisms of sleep-mediated memory consolidation and schema-driven distortions in remembered durations.

In summary, the results of this study emphasize the essential role of retrieval and memory consolidation in the interaction between schematic knowledge and episodic memories. The experiment described in Chapter 4 is to my knowledge the first study to demonstrate the role of sleep-mediated memory consolidation in language-driven memory distortion. These findings are consistent with prior schema-based memory theories in that prior knowledge brought to mind to interpret an event are integrated with episodic details during consolidation, as discussed in Chapter 1. However, we have additionally shown that the presence of language at encoding can modulate memories in a similar way to conceptual schemas (e.g., prior knowledge about restaurants). This not only introduces a novel way to manipulate prior knowledge but also illuminates the debate on the relationship between language and memory. Language-induced conceptualizations do not appear to play a role until after some consolidation has occurred, unless language cues recollection. This was observed in duration reproductions and in verbal recall, suggesting that several aspects of event representation were integrated with verbal concepts, e.g., the way geometric figures are referred to and likely represented. These results therefore suggest that the initial verbal encoding of a memory is likely to bias how events are recalled after some period of consolidation, as verbal concepts are merged with episodic details. This finding is inconsistent with interactive encoding accounts of language and memory as well as retrieval accounts. Instead, they suggest a new possibility: that language shapes event memory during memory consolidation, even when language does not cue retrieval. Although generally consistent with schema-based memory theories, these results are novel with respect to the specific role of language in memory.

6.3.4 Contributions to theories of the role of sleep in memory

An interesting finding in chapter 4 was that duration reproductions became longer after sleep, as did verbal recallections. We interpret these findings as indicating that conceptual and episodic features were merged, and thus there was an increase the number of associations brought to mind during retrieval after sleep. This finding thus suggests that sleep increases rather than reduces the number of details re-activated, and it is thus inconsistent with proposal suggesting that only gist-like information is retained due to decay, as discussed in Chapter 1. Instead, this finding suggests that memory stabilization is indeed achieved via initial enrichment of episodic memories by conceptual knowledge taking place during sleep. This possibility adds a new way of thinking about the processes thought to take place during sleep (Walker & Stickgold, 2013). Integration during sleep does not imply loss of episodic details, instead it implies enrichment, even though with more time, episodic details may indeed be lost (e.g., a week later). This therefore suggests a different development for memories over time previously unsuspected: mental associations increase after sleep, with decay only starting after that.

However, how sleep-mediated memory consolidation contributes to schema-driven distortions in reproduced duration and event memory more generally remains unclear. It would be necessary to carry out further experiments to explore this question more deeply. It is possible for example, that the effects we have found are specific to verbal encoding, rather than to event schemas more generally. In addition, the findings presented and discussed in Chapters 3 and 4 also suggest the possibility that the nature of the stimuli which were used (ambiguous speed, and the mixing of goal-specific and goal-unspecific events) modulated the initial mapping between linguistic concept and episodic details at the encoding stage. Further study is needed to test this idea.

6.4 Putting things together / conclusions

The findings presented in this thesis have increased our understanding of how human beings remember event duration and how they represent it in terms of language comprehension. I have suggested two models to explain recollection of real-world experience and information processing in language comprehension: the recall-based model and the activation-based model. In both models, it is suggested that information activation is central for memory recollection and language understanding of event duration, and time, more generally. Specifically, the representation of event duration is determined by both the quantity (the amount of recalled information) and the quality (the diversity of recalled information and the concreteness of recalled information) of the activated information. In language comprehension, the information activated during processing is not only lexical in nature but is also derived from schematic knowledge. Similarly, memory recall for a specific event duration is driven by the activation of initially encoded episodic details and the associated schematic information. In both cases, longer processing time is expected if there is more diverse knowledge or a larger amount of information activated during the recall process. Hence, duration representations are grounded in the information activated in each circumstance.

Taking these findings together, I suggest that information activation is the basis of event duration representation in language and real-world experiences. In contrast to previous theories, I have pinpointed the influence of the quantity or quality of activated information in event duration representation and proposed an information-activation based model to explain memory and comprehension of event duration. I have also highlighted the role of retrieval cues in shaping memory of event duration. When language serves as a retrieval cue, event schemas interact with episodic event details and lead to biases of episodic features towards the conceptual features which are contained in event schemas.

APPENDICES

Appendix 1 Supplemental Materials (Chapter 2)

Appendix 1.1 Stories of large timescale (beyond one day)_Expt 1, Expt2, Expt3

NO.	Large Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
1	Dave went to Liverpool for a job. His contract was for a few months/a few weeks. It was just enough time for him to assemble the boat. He was happy that he got a lot of money.	Liverpool	assemble	Was Dave's contract for a few weeks?
2	Richard's company moved him to Ecuador. He worked very efficiently there for a few months/a few weeks. He spent all that time building a house. He was exhausted afterwards.	Ecuador	house	Did Richard build a house in Ecuador?
3	Mike's new office required a lot of work. A month/Two weeks went by and he could not move in It took all that time to change the pipes Services are not great around here	Mike's office	pipes	Did it take a month to change the pipes?
4	Doug sold his flat and moved out. His new place was not ready until a month later/a week later. It took him that long to change the roof tiles. In the end, he broke even.	flat	roof	Did it take a month to change the tiles?
5	Bob's mum is a bit overbearing. For example, last Christmas she visited him for a few weeks/a few days. She spent the whole time cleaning the house. Bob did not have a good time.	overbearing	cleaning	Did Bob's mother visit around Christmas time?

No.	Large Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
6	The council approved the plans for a new industrial estate. but the building work was delayed for six months/a month. It took them that long to cut down the forest. Opinions on the news were very critical.	council	forest	Was the forest cut down in a week?
7	Arthur loved his work. Once, he secluded himself in his studio for a month/a week. He devoted all that time to carving a statue. Afterwards, he was very pleased.	Arthur	carving	Was Arthur carving for a week?
8	Lucy's parents went on holiday. They drove a campervan for a couple of months/a couple of weeks. It took them all that time to cross America. Back in the UK, they had a big party.	holiday	America	Were Lucy's parents on holiday?
9	Several prisoners had escaped and the government was furious. The prison director was given one month/two weeks to improve security. It was just enough time to erect a wall. The press was all over the incident.	government	erect	Was a wall erected in a week?
10	Tim worked for an engineering company. He was stationed in Russia for a month/a week. It took him all that time to refurbish an oil rig. Later, the company congratulated him for the job.	Tim	rig	Was Tim stationed in Russia?

No.	Large Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
11	John was supposed to be in Aberdeen by now. but he was delayed for a week/ a couple of days due to extra work . He needed all that time to unload a cargo train. In the way back, he decided to quit his job.	Aberdeen	unload	Was the train unloaded in a week?
12	Jenny got very depressed after her boyfriend joined the army. Because of this she went to her room and didn't come out for two weeks/a week. She spent all that time knitting a sweater. When he boyfriend came back, she was delighted.	boyfriend	knitting	Did Jenny's boyfriend join the foreign secret service?
13	Ted worked for the Red Cross. His assignment in Africa was delayed for two weeks/three days. It took that long to fill up his cargo plane. There had been a problem somewhere.	Red Cross	plane	Was Ted delayed for a few hours?
14	John wanted to surprise his relatives for Christmas. He took a month off work/two weeks off work. It was just enough time for him to paint the house. He did not get his salary but he was happy.	relatives	house	Did John extend the house this year?
15	The manager ordered the evacuation. The employees were relocated for a month/a week. It took all that time to thoroughly fumigate the building. The company's work was seriously disrupted.	evacuation	building	Was the work disrupted during the relocation?

No.	Large Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
16	Jim was arrested for stealing. He spent a few weeks/ a few days in prison. It took him that long to read "The Da Vinci Code". Afterwards, he watched the movie.	stealing	read	Did Jim read a book in a few weeks?
17	Phil was nervous about the competition. He had two weeks/a week to prepare. It was barely enough time for him to repair his motorbike. His friends were happy for him.	competition	motorbike	Was Phil nervous?
18	Jessica strained her ankle. She couldn't walk for a week/a few days. It was just enough time for her to sew her prom dress. Her mother was happy that she found something to do.	ankle	sew	Did Jessica sew her wedding dress?
19	The mayor wanted to improve the look of the city. He had two months/one month before the election. It took that long to construct the bridge. As expected, the election was a success.	mayor	bridge	Was the bridge built before the election?
20	The city council noticed the pests in June. They gave Terry a deadline of two weeks/one week to deal with the problem. It took him that long to thoroughly spray the park. The neighborhood was very grateful.	council	spray	Did Terry work on the park for a day?

No.	Large Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
21	Alice finally had a chance to use her skills. She got a two-months/one-month contract with a French company. It took her that long to translate a book. Later, she became a permanent employee.	skills	book	Was Alice working for a German company?
22	John lost his job in May. For the next four months/two months, he decided to fulfill his dream. He spent all that time writing his novel. His girlfriend broke up with him though.	lost	writing	Did John write a play in a few months?
23	The company is expanding. Jonathan was sent to Canada for two months/ two weeks. He spent the whole time drilling for oil. He had an exciting time there.	company	drilling	Was Jonathan working for the oil industry?
24	Paul is always called out to solve difficult problems. Once he was gone for a month/a week. He needed all that time to restore a plane. He is very knowledgeable.	problem	restore	Did Paul restore a plane in a month?
25	Rick was a policeman, but he was not very good at his job. Once he was travelling on the job for two months/a month. It took him all that time to arrest a thief. The local press commented on his skills.	not good	thief	Did it take a week for Dick to arrest the thief?

No.	Large Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
26	Betty works in a big hotel. She's been on duty for the last week/last two days. It took her all that time to iron the sheets. And her co-workers are not much help.	hotel	sheets	Did Betty spend her shift at the hotel ironing sheets?
27	Bill was worried about the damage the storm had caused to his land. He was gone from the house working for a month/a week. He needed all that time to fix his fence. He did not have any help.	storm	fence	Did it take Bill a week to fix the fence?
28	Jenny had always been adventurous. After finishing her degree, Jenny went to Canada for a month/two weeks. She spent the whole time panning for gold. She had always wanted to do that.	adventurous	panning	Did Jenny spend a week panning for silver?

Appendix 1.2 Stories of small timescale (within one day)_ Expt 1, Expt 2

No.	Small Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
1	Lisa was moving to a new flat near the university. John wanted to help and spent his morning/an hour there. It took him all that time to assemble her bed. He was really tired afterwards.	flat	bed	Did John help Lisa?
2	Peter rushed to work this morning. He got there half an hour/ ten minutes late. It took him that long to sew on his button. He was a bit annoyed.	work	sew	Was Peter half an hour late for work?
3	Tom and Beth were in Rome on holiday. Beth took off for the afternoon/an hour to go shopping. Tom spent all that time finding an old church. They met up for dinner later that night.	Rome	church	Were Beth and Tom in Rome because of work?
4	Ms Wilson didn't have much to do. So she sat on the sofa for a few hours/an hour. She devoted that time to knitting a coaster. Later, she made dinner.	Wilson	knitting	Did Mrs Wilson fall asleep on the sofa?
5	Ms Smith enjoyed spending time with her friends. Yesterday, she sat down with them for the entire morning/ a couple of hours. She spent the whole time embroidering a cushion. She forgot about her appointment.	friends	embroidering	Did Ms Smith play cards with her friends?
6	Alice was in the garden weeding the flowerbeds. She took a break for a couple of hours/ half an hour. She spent all that time washing her dog. She got all wet.	weeding	washing	Did Alice wash the dog in a couple of hours?

No.	Small Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
7	Molly felt like doing some work around the house. She had three hours/ an hour to spare before her appointment. She spent all that time altering her dress. She then left for her appointment.	work	dress	Did Molly take a nap before her appointment?
8	Ben was getting ready for a big paint-ball game with his friends. He went into the woods for several hours/ an hour. He spent all that time digging a hole. He was planning a trick.	game	digging	Did Ben dig a hole for that paintball game?
9	Vicky bought a mobile phone in Germany. She couldn't operate it for a day/ a few hours. It took her that long to translate the instructions. She was a bit frustrated.	Germany	translate	Did Vicky manage to use her phone straight away after she bought it?
10	Tony and Natalie broke up. She went to her room for an hour/a few minutes. It took her that long to read his letter. She was devastated.	broke	letter	Did Tony go to his room for a while?
11	Jessie had an exam on Monday. She started writing half an hour/a few minutes late. It took her that long to fix her pen. She was very anxious.	exam	pen	Did Jessie begin her exam half an hour late?
12	When Ken was little, he was a troublemaker. Once, he disappeared from school for five hours/two hours. It was just enough time for him to climb a pylon. He was very proud of himself	trouble	pylon	Was Ken an obedient child?
13	After she came back from work, Claire was in a horrible mood. She sat at her desk for about an hour/a few minutes. She spent all that time writing her resignation. She had fought with her boss.	mood	resignation	Did Claire write a letter in a few hours?

No.	Small Timescale Story (long VS. short)	Early probe	Late Probe	Comprehension question
14	There was a big accident on the highway today. The traffic was chaotic for about an hour/fifteen minutes. It took that much time for the police to erect the roadblocks. Tempers were high.	accident	roadblocks	Did it take an hour to erect the roadblocks?
15	Ms Smith had many chores to do around the house. Before going to work, she stayed at home for a couple of hours/an hour. It took her that long to spray the roses. By this time, she was quite tired.	chores	spray	Did Ms Smith stay at home before work?
16	When little Kevin was on holiday with his parents, he had to be watched all the time. Only once did he play by himself for the whole afternoon/ the whole hour. He spent all of that time building a sandcastle. The parents could relax a bit then.	holiday	sandcastle	Did Kevin build a sandcastle in a few minutes?
17	Mike's mother was late picking him up from school. He had to wait for about an hour/a quarter of an hour. He devoted that time to designing a spaceship. A friend was with him.	late	spaceship	Did Mike design a spaceship?
18	Tom is a street artist. Yesterday, he sat there for the whole afternoon/ an hour. It took him that long to draw a tiger. Passers-by admired his work.	artist	tiger	Did Tom spend the whole afternoon drawing a tiger?
19	Steve wanted to change the view from his window. He went out to work in the garden for several hours/an hour. It took him all that time to chop down a tree. Now, he had to get rid of the stump.	view	tree	Was Steve working in the garden for about an hour?
20	Rebecca went to visit her sister. She stayed for a couple of hours/ for the afternoon. They spent that time baking a cake. Rebecca went to the movies afterwards.	visit	cake	Did Rebecca watch the TV during her visit at her sister's?

A1.3 Stories of state story_ Expt 3

No.	State Story (long VS. short)	Early probe	Late Probe	Comprehension question
1	Mary was always a difficult daughter. Once, she had a huge argument with her mother about her boyfriend. For several weeks/ a few weeks, she had not realized her boyfriend was a drug addict. In the end, she admitted her mistake.	difficult	realized	Did Mary have an argument with her farther?
2	Jill was disappointed with herself. She told her friends about it. She had been on a diet for two months/a month. But she had not managed to lose a single pound.	herself	diet	Did Jill tell her parents her feelings?
3	John loved his dog and had it for years. When the dog died, he took it badly. He was depressed for months/weeks. In the end, he bought a cat to feel better.	loved	depressed	Was John depressed for weeks?
4	When Helen was a child, she lived in a big house with her mother. One night, there was a huge storm and blackout. Afterwards, she was afraid of storms for years/months. When she grew older, she conquered her fear.	mother	afraid	Was Helen afraid of storms for all her life?
5	Alice studied Spanish at university. She spent her year abroad in South America. When she was there, she was homesick for a few months/ a few weeks. But afterwards, she got used to her new life.	Spanish	homesick	Did Alice study in a language school?
6	Chris has a flower shop in the city. This year, he was very unlucky with the floods. His shop had to be closed for two weeks/a week. But, he managed to make up for the loss.	flower	closed	Was the shop closed for a week?

No.	State Story (long VS. short)	Early probe	Late Probe	Comprehension question
7	Laura has two children and lives away from her parents. Lately, she has started to worry about her absence from work. Her children had been ill for about a month/ two weeks. And her boss is very strict.	parents	ill	Did Laura and her children live with her parents?
8	Dan is from Asia but he studied in Manchester. At the beginning, it was difficult for him to adjust to his new life. He had no friends for a several months/a few months. But then he found a way to connect with people.	Asia	friend	Was Dan still isolated in the new city?
9	Jeremy works as a teacher. He loves his job but it is not well paid. He has lived with a colleague for the last two years/last year. But now, he wants his own house.	works	colleague	Did Jeremy live with his colleague for one year?
10	Kevin moved in with his girlfriend. But there was a problem. They did not have any heating for two weeks/a week. Finally, the landlord repaired the boiler.	girlfriend	heating	Was Chris happy with his flatmates?
11	Sarah and Ben got married this spring. During their honeymoon, Ben hit his head in a motorcycle accident. He was in a coma for two weeks/a week. Sarah was so worried about him.	spring	coma	Did Sarah have an accident?
12	Chris' flat mates like to play pranks on each other. Once they called his girlfriend pretending to be him but Chris did not take it well. He was angry at them for months/for weeks. Eventually, he moved out of the house.	pranks	angry	Was Chris happy with his flatmates?

No.	State Story (long VS. short)	Early probe	Late Probe	Comprehension question
13	Courtney went to visit her brother in Nevada. She had a car accident and got injured. She was not able to walk for a couple of months/weeks. She had damaged her knees.	Nevada	walk	Did Courtney visit her brother?
14	Tom was always very active and energetic. Last year, he felt unwell, so he went to the to see his doctor. He was on medication for months/weeks. He decided to be more careful.	enegetic	medication	Was Tom on treatment for months?
15	Susan's neighbour had been burgled. They took all of their valuables and threatened the family. She was nervous about it for weeks/days. In the end, she installed a burglar alarm.	neighbour	nervous	Did Susan install a burglar alarm?
16	Julia wanted to try something different. So she started golfing. She felt insecure about it for months/weeks. Afterwards, she started to feel better about it.	wanted	insecure	Did Julia begin to learn golfing?
17	Beth and Eleanor were really good friends. They knew each other since childhood. And they were neighbours for two years/a year. They also called each other all the time.	Eleanor	neighbours	Were Beth and Eleanor neigbours for two years?
18	Dr. Jones had always been very active. But he fell down the stairs. He was in a wheelchair for two months/a month. Then, his condition improved very quickly.	active	wheelchair	Did Dr.Jones fall down the stairs?
19	Joe loved to ride his bike. He once cycled from London to Paris. His legs hurt for weeks/ days. He was happy with his achievement.	loved	hurt	Did Joe's legs hurt for weeks?

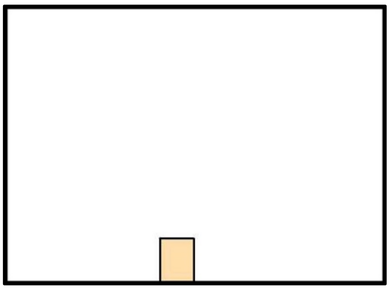
No.	State Story (long VS. short)	Early probe	Late Probe	Comprehension question
20	Tim and Lisa were classmates at university. When they got their first jobs in London, they decided to live together. They owned a flat for two years/a year. Then, they moved to different cities.	university	flat	Did Tim and Lisa lived together in London for two years?
21	Emily had to change schools last year. There were some really strange characters in her class. One girl had blue hair for a few months/a few weeks. That was when the teachers decided to tighten to rules.	change	blue hair	Did Emily change her school?
22	Harry loved the Olympics. He watched his favourite gymnast in London. He was angry at the judges for days/hours. They had made very questionable decisions.	Harry	angry	Did Harry watch the gymnast in London?
23	Simon worked hard as a lawyer in New York. But he enjoyed his spare time. He owned a weekend house for three years/ a year. Until the commute was too tiring.	New York	weekend house	Was Simon a hardworking lawyer?
24	Tony's dog would not eat or move. Beth recommended the neighborhood vet. She had known him for two years/a year. And he had a good reputation.	eat	known	Had Beth known the vet for two years?

Appendix 2: Supplemental Materials (Chapter 3 & Chapter 4)

Appendix 2.1 Stimulus phrases and cue frames

A firework rocket launched into the sky

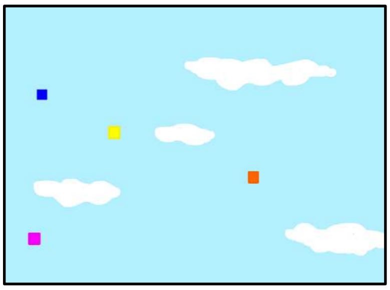
A Chinese lantern raising up into the sky



(3 sec., 2.2 segments)

Planes flying in the sky

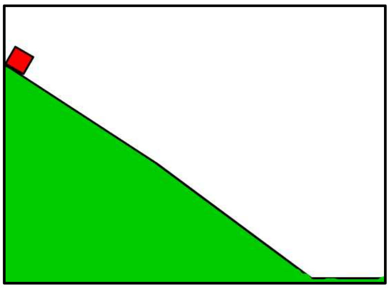
People paragliding in the sky



(3 sec., 3.3 segments)

A racing car going downhill.

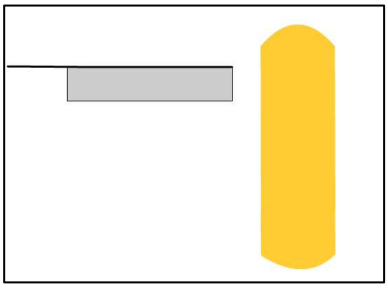
An army tank going downhill



(3 sec., 2.2 segments)

Cutting a banana with a knife

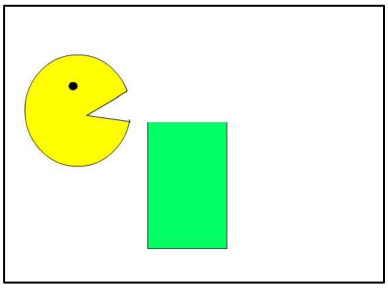
Cutting a log with a handsaw



(4 sec., 4.1 segments)

Someone gulping a glass of beer.

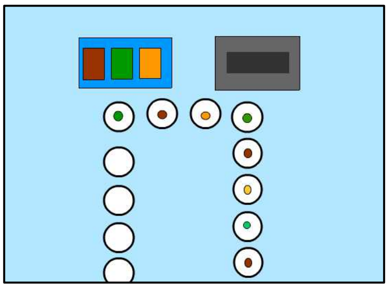
Someone drinking a glass of beer



(4 sec., 3.2 segments)

Products in a factory's assembly line

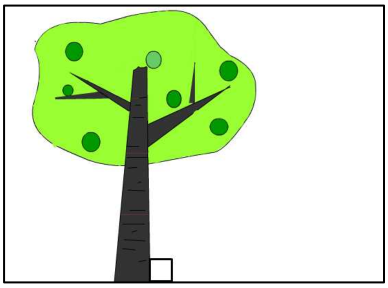
People lining up to buy food in a canteen



(4 sec., 3.5 segments)

A monkey climbing a tree to pick a fruit

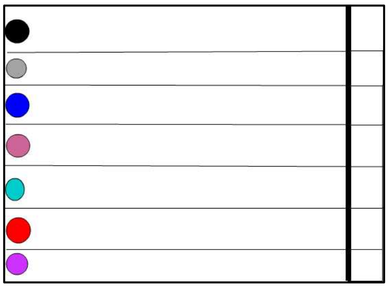
An old man climbing a tree to pick a fruit



(5 sec., 3.6 segments)

Children racing in a school yard

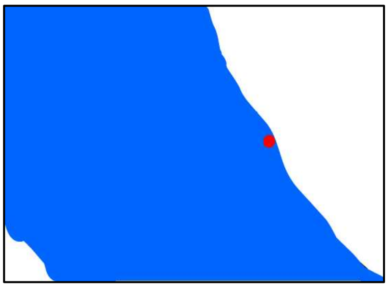
Horses sprinting ar a race course



(5 sec., 7.1 segments)

A speed boat crossing a river

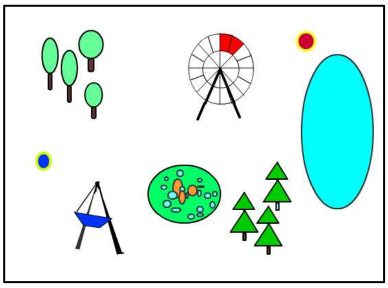
A man swimming across a river



(5 sec., 5.9 segments)

Two friends jogging to meet in the park

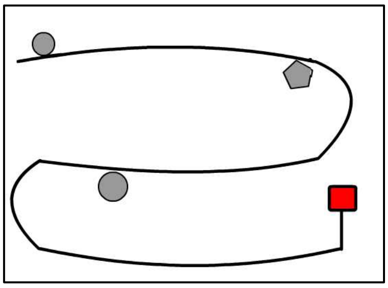
Two friends strolling to meet in the park



(6 sec., 4.4 segments)

Cyclists in a relay competition

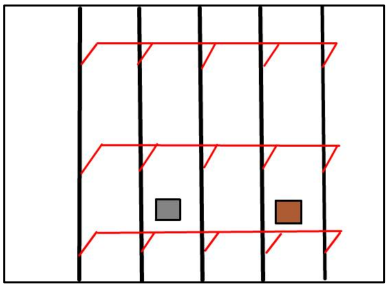
Students in a relay competition



(6 sec, 4 segments)

Athletes competing in a hurdle race

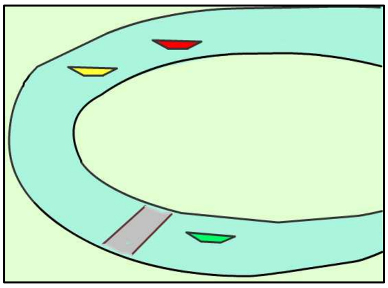
Army cadets crawling under wire obstacles



(6 sec, 5.7 segments)

Kayakers competing in the river

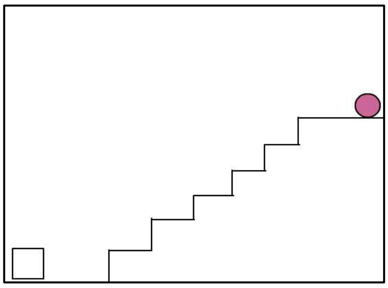
Visitors rowing in the river



(7 sec., 2.6 segments)

A child running downstairs to find mom

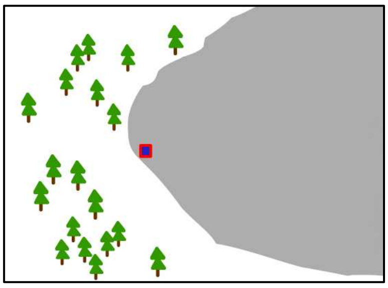
A child crawling downstairs to meet mom



(7 sec., 6.6 segments)

A person skating on a frozen lake

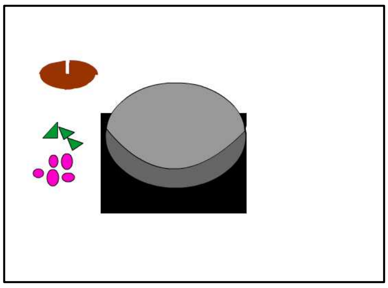
A person stepping on to a frozen lake.



(7 sec., 2.9 segments)

Mixing ingredients in the bowl

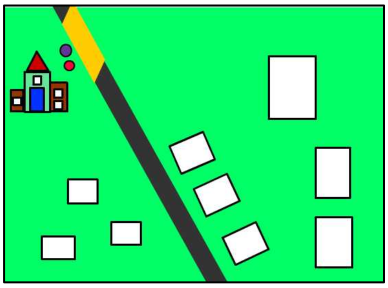
Simmering vegetables in the pot



(8 sec., 5 segments)

Taking the tube home

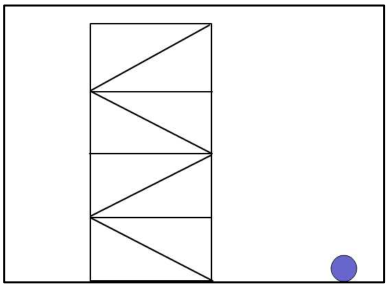
Taking the school bus home



(8 sec., 4.1 segments)

Students running up to their classroom

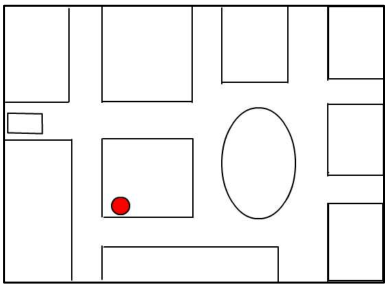
Students walking up to their classroom



(8 secs, 5.7 segments)

An ambulance taking someone to the hospital.

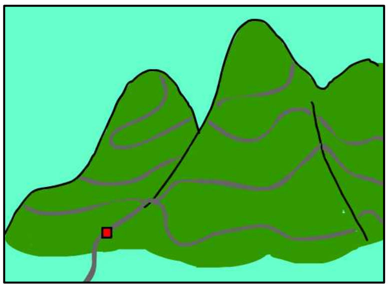
Grandma taking the bus to the hospital



(9 sec., 6.2 segments)

A tourist driving a car on a mountain road

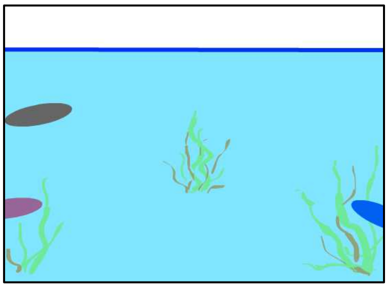
A tourist riding a mule on a mountain path



(9 sec., 2.4 segments)

Dolphins swimming in an aquarium pool

Goldfish swimming in an aquarium tank



(9 sec., 3.8 segments)

- **Appendix 2.2 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 3.1. 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 3.1. 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 3.1 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 3.1. 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.

● Appendix 2.3. 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 A1 for the three animation. 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.

Table A1: Example of verbal recall in Experiment 1A with coded changes and number of words

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

● Appendix 2.4 Full models from main experiment of Chapter 4 (Sleep study)

Table A2: results from the full model of the main experiment

RD	Estimate	SE	df	t	P
(Intercept)	3600.51	297.08	24.00	12.12	0.00
Language condition	-18.46	85.88	18.00	-0.22	0.83
Imm. vs Delay	513.99	77.13	181.00	6.66	0.00
Wake vs Sleep	-236.86	132.88	182.00	-1.78	0.08
Duration	359.69	47.28	25.00	7.61	0.00
Language: Imm. vs Delay	102.47	50.32	3387.00	2.04	0.04
Language: Wake vs Sleep	-125.52	86.72	3380.00	-1.45	0.15
Language: Duration	7.64	13.61	18.00	0.56	0.58
Imm. vs Delay: Duration	-56.77	12.78	181.00	-4.44	0.00
Wake vs Sleep: Duration	-24.92	22.03	181.00	-1.13	0.26
Language: Imm. vs Delay: Duration	-14.72	7.98	3388.00	-1.85	0.07
Language: Wake vs Sleep: Duration	11.36	13.75	3382.00	0.83	0.41
Deviation Index					
(Intercept)	1.75	0.06	35.00	27.62	0.00
Language	0.01	0.02	18.00	0.67	0.51
Imm. vs Delay	0.15	0.02	185.00	6.17	0.00
Wake vs Sleep	-0.13	0.04	185.00	-3.01	0.00
Duration	-0.12	0.01	26.00	-12.80	0.00
Language: Imm. vs Delay	0.02	0.01	3374.00	2.44	0.01
Language: Wake vs Sleep	-0.04	0.02	3367.00	-2.29	0.02
Language: Duration	0.00	0.00	18.00	-0.39	0.70
Imm. vs Delay: Duration	-0.02	0.00	182.00	-6.60	0.00
group2: Duration	0.01	0.00	182.00	1.96	0.05
Language: Imm. vs Delay: Duration	0.00	0.00	3383.00	-2.13	0.03
Language: Wake vs Sleep: Duration	0.00	0.00	3377.00	1.64	0.10

● Appendix 2.5 Replication of language effect after sleep

Participants 90 participants who reported being English native speakers (45 participants in list1 and 45 participants in list2) were collected from Prolific.ac and the University of York (60 participants were from prolific, 30 participants from the University of York). Participants were awarded course credits or a small payment. The aim was to collect as many participants as possible. Since responses on Prolific were not sufficient, we contacted university first year students (who have just arrived at York) via email.

Design and procedure The design and procedure were similar to that of the Sleep group in our main experiment but they differ in two main respects: (1) reproduced durations were measured using an online site (https://www.testable.org/) and a key press, rather than a mouse click—participants pressed the space bar to start and finish the mental replay— and (2) the verbal recall task was replaced by a phrase recall task. This task allowed us to establish whether participants learned something about the animations at the same time that made the experiment shorter. Writing down all details about the animations was time consuming and risked participants abandoning the experiment. Thus, from this phrase recall task, we simply check that participants provided some information that matched the animation, regardless of whether the phrase was exactly correct or not. When participants registered on the Prolific website or the university recruitment site, they were provided with the link to the learning phase of the study, together with the general description of the study and consent form. They were informed about the time requirements to carry out each of the two component sessions of the study. In particular, they were required to start the learning session in the evening. At the end of the learning session, each participant emailed the experimenter to receive a link to the test session. The web page controlling the experiment recorded the time at which each participant started each session. If participants attempted to start the test session before 10 hrs. have passed, the web page would not proceed. The average start time of the learning session was in the evening (between 18:43 and 0:44) and the test session was completed the next morning (between 7:13 and 13:27). The average time between the two sessions was around 12.40hs.

Data treatment The main dependent variables were the reproduced duration in milliseconds and the deviation index. Due to the uncontrolled nature of the testing, we expected that some participants would not complete the study (e.g., would not complete the second session) or would quickly press the key throughout the mental replay task without actually replaying the animation. 12 participants who did not provide any phrase recollection data or pressed the key at intervals smaller than a second in the replay task were excluded from the analyses. A total of 78 participants were then analysed (39 in list 1 and 39 in list 2).

We removed outliers falling beyond 1.5 x the group's interquartile ranges (IQR) and those falling below 1s. We also removed trials for which participants did not provide any recall information. These exclusions comprised less that 4% of the data set. Hypotheses testing used linear mixed-effects models including the maximal random effects structure allowed by convergence (i.e., random intercepts for subjects and items, by-subjects random slopes for stimulus duration and by-item random slopes for language condition) (Barr et al., 2013).

Results

Table A3 shows the results of this study for reproduced duration and deviation indices (see also Figure A1). There was a main effect of condition and duration for reproduction duration, indicating a correlation with stimulus duration and an effect of language. There was also an effect of language for deviation indices indicating over-reproductions for animations described as slow. Despite the uncontrolled nature of the testing and the change in response measure, this web-based study replicates the main finding of our lab-based study.

Table A3: results from the full model of the replication of sleep experiment

DV	Fixed effects	Estimate	SE	df	t	p
Reproduced duration	(Intercept)	3763.83	291.37	26.1	12.92	<0.001
	Duration	314.67	47.97	30.28	6.56	<0.001
	Language condition	92.73	40.07	60.64	2.314	<0.02
Deviation Index	(Intercept)	1.02	0.07	29.14	15.42	<0.001
	Language Condition	0.02	0.02	43.97	2.43	0.02

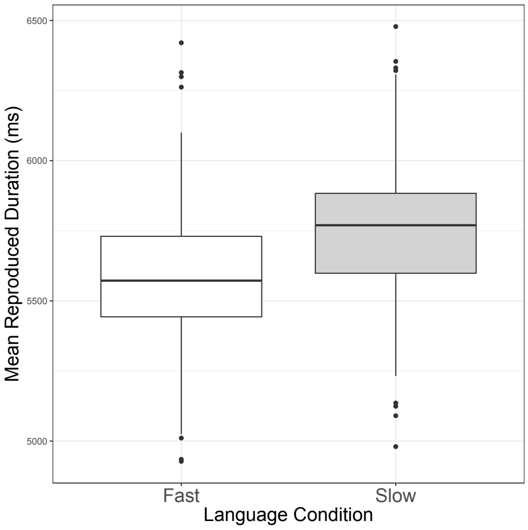


Figure A1: Mean reproduced durations across participants. Horizonal lines indicate the medians of each distribution.

REFERENCES

Abelson, R. P. (1981). Psychological status of the script concept. American psychologist, 36(7), 715.

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

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. Behavioral and brain sciences, 22(03), 425-444.

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

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

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. Cognitive science, 33(4), 583-609.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. Journal of educational psychology,51(5), 267.

Avni-Babad, D., & Ritov, I. (2003). Routine and the perception of time. Journal of Experimental Psychology: General, 132(4), 543.

Ashton, J. E., Jefferies, E., & Gaskell, M. G. (2018). A role for consolidation in cross-modal category learning. *Neuropsychologia*, *108*, 50-60.

Baddeley, A., Eysenck, Michael W, & Anderson, Mike. (2009). Memory. Hove: Psychology Press.

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

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. Child development, 72(3), 708-717.

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

Barsalou, L. W. (1999). Perceptions of perceptual symbols. Behavioral and brain sciences, 22(04), 637-660.

Bartlett, F. C. (1932). Remembering. Cambridge. UK: Cambridge UP.

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

Berg, M. (1979). Temporal duration as a function of information processing. Perceptual and motor skills.

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. (1989). Experiencing and remembering time: Affordances, context, and cognition. Advances in psychology, 59, 333-363.

Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In Time, action and cognition (pp. 141-152). Springer Netherlands.

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. (1996). Models of psychological time revisited.Time and mind, 33, 171-195.

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

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. Acta psychologica, 134(3), 330-343.

Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgments: A meta-analytic review. Psychology and Aging, 13(4), 584.

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. (1992). The remembering of auditory event durations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(5), 938.

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

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of verbal learning and verbal behavior, 11(6), 717-726.

Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. Academic.

Brown, S. W. (2008). Time and attention: Review of the literature. Psychology of time, 111-138.

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

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

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. Psychiatry Research, 28, 193–213.

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. Journal of Experimental Psychology, 15(1), 73.

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. Memory & Cognition, 25(4), 438-446.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive psychology, 4(1), 55-81.

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), 41-79.

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(5), 824.

Dumay, N., & Gaskell, M. C. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. Psychological Science, 18(1), 35–39.

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

Fecica, A. M., & O'Neill, D. K. (2010). A step at a time: Preliterate children's simulation of narrative movement during story comprehension. Cognition, 116(3), 368-381.

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 concepts. Journal of Memory and Language, 44(4), 516-547.

Fraisse, P. (1963). The psychology of time.

Friedman, W. J., & Janssen, S. M. (2010). Aging and the speed of time. Acta Psychologica, 134(2), 130-141.

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.

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. Cognition, 89(1), B27-B41.

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.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. Neuropsychologia, 53, 104-114.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. Annals of the New York Academy of sciences, 423(1), 52-77.

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: advances and challenges. Current opinion in neurobiology, 7(2), 170-184.

Grady, C. L., Springer, M. V., Hongwanishkul, D., McIntosh, A. R., & Winocur, G. (2006). Age-related changes in brain activity across the adult lifespan. Journal of cognitive neuroscience, 18(2), 227-241.

Grafman, J. (1995). Similarities and distinctions among current models of prefrontal cortical functions. Annals of the New York Academy of Sciences, 769(1), 337-368.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. Psychological bulletin, 127(1), 22.

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. Attention, Perception, & Psychophysics, 72(3), 561–582.

Gu, B. M., & Meck, W. H. (2011). New perspectives on Vierordt’s law: memory-mixing in ordinal temporal comparison tasks. In Multidisciplinary aspects of time and time perception (pp. 67-78). Springer, Berlin, Heidelberg.

Guyau, J. M. (1890). La genèse de l'idée de temps. Alcan.

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

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. Memory & cognition, 34(6), 1221-1235.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. Neuron, 41(2), 301-307.

Hennies, N., Lambon Ralph, M. A., Kempkes, M., Cousins, J. N., & Lewis, P. A. (2016). Sleep Spindle Density Predicts the Effect of Prior Knowledge on Memory Consolidation. The Journal of Neuroscience, 36(13), 3799–3810. https://doi.org/10.1523/JNEUROSCI.3162-15.2016

Hicks, R. E., Miller, G. W., & Kinsbourne, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. The American journal of psychology, 719-730.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: a new approach. Psychophysiology, 10, 431–436.

Hoerl, C., & McCormack, T. (2001). Time and memory: Issues in philosophy and psychology (No. 1). Oxford University Press.

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. Behavior research methods, 45(3), 718-730.

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. International Journal of Chronobiology, 4, 97–110.

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

James, W. (1890). The principles of psychology, Vol. 2. NY, US: Henry Holt and Company.

Janssen, S. M., Naka, M., & Friedman, W. J. (2013). Why does life appear to speed up as people get older?. Time & Society, 22(2), 274-290.

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. (2018). Temporal compression in episodic memory for real-life events. Memory, 26(6), 759-770.

John, D., & Lang, F. R. (2015). Subjective acceleration of time experience in everyday life across adulthood. Developmental Psychology, 51(12), 1824.

Kapur, N. (1999). Syndromes of retrograde amnesia: a conceptual and empirical synthesis. Psychological bulletin, 125(6), 800.

Kelter, S., Kaup, B., & Claus, B. (2004). Representing a described sequence of events: a dynamic view of narrative comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(2), 451.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. Trends in cognitive sciences, 12(2), 72-79.

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. Memory & cognition, 39(1), 75-91.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP).Annual review of psychology, 62, 621.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203-205.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13). https://doi.org/10.18637/jss.v082.i13

Landmann, N., Kuhn, M., Piosczyk, H., Feige, B., Baglioni, C., Spiegelhalder, K., … Nissen, C. (2014). The reorganisation of memory during sleep. Sleep Medicine Reviews, 18(6), 531–541. https://doi.org/10.1016/j.smrv.2014.03.005

Lejeune, H., & Wearden, J. H. (2009). Vierordt's The Experimental Study of the Time Sense (1868) and its legacy. European Journal of Cognitive Psychology, 21(6), 941-960.

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. Psychology and aging, 17(4), 677.

Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. Cognitive Psychology, 12(3), 412-445.

Lindsay, S., Scheepers, C., & Kamide, Y. (2013). To dash or to dawdle: Verb-associated speed of motion influences eye movements during spoken sentence comprehension. PloS one, 8(6), e67187.

Loftus, E. F. (1975). Leading questions and the eyewitness report. Cognitive psychology, 7(4), 560-572.

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., & Meltzoff, A. N. (2013). Goals influence memory and imitation for dynamic human action in 36-month-old children. Scandinavian journal of psychology, 54(1), 41-50.

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

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological review, 101(4), 676.

Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus intensity and the perception of duration. Journal of Experimental Psychology: Human perception and performance, 37(1), 303.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition*, *32*(8), 1389-1400.

Mcclain, L. (1983). Interval estimation: Effect of processing demands on prospective and retrospective reports. Perception & Psychophysics, 34(2), 185-189.

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

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. Journal of memory and language, 66(4), 545-567.

Mioni, G., Zakay, D., & Grondin, S. (2015). Faster is briefer: The symbolic meaning of speed influences time perception. Psychonomic bulletin & review, 22(5), 1285-1291.

Moscatelli, A., & Lacquaniti, F. (2011). The weight of time: gravitational force enhances discrimination of visual motion duration. Journal of Vision, 11(4), 5-5.

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

Mulligan, R. M., & Schiffman, H. R. (1979). Temporal experience as a function of organization in memory. Bulletin of the Psychonomic Society, 14(6), 417-420.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesi

Neuschatz, J. S., Lampinen, J. M., Preston, E. L., Hawkins, E. R., & Toglia, M. P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. Applied Cognitive Psychology. https://doi.org/10.1002/acp.824

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. Journal of Personality and Social Psychology, 28(1), 28.

Newtson, D. (1976). Foundations of attribution: The perception of ongoing behavior. New directions in attribution research, 1, 223-247.

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

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: a memory-based text processing view. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(5), 1200.

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

Ornstein, R. E. (1975). On the experience of time.

Palmer, S. E. (1999). Vision science: Photons to phenomenology. MIT press.

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. Memory & Cognition, 3, 519-526.

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

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L. W., Wamsley, E. J., Tucker, M. A., … Stickgold, R. (2009). The role of sleep in false memory formation. Neurobiology of Learning and Memory, 92(3), 327–334. https://doi.org/10.1016/j.nlm.2009.03.007

Penney, T. B., & Vaitilingam, L. (2008). Imaging time. Psychology of time, 261-294.

Poynter, D. (1989). Judging the duration of time intervals: A process of remembering segments of experience. Advances in psychology, 59, 305-331.

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.

Predebon, J. (1984). Organization of stimulus events and remembered apparent duration. Australian journal of psychology, 36(2), 161-169.

Predebon, J. (1996). The effects of active and passive processing of interval events on prospective and retrospective time estimates. Acta Psychologica, 94(1), 41-58.

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., & 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., 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., 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.

Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. The projection of arguments: Lexical and compositional factors, 97-134.

Rasch, B. H., & Born, J. (2013). About sleep’s role in memory. Physiological Reviews, 96(2), 681–766. https://doi.org/10.1152/Physrev.00032.2012

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

Rempel-Clower, N. L., Zola, S. M., Squire, L. R., & Amaral, D. G. (1996). Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. The Journal of Neuroscience, 16(16), 5233-5255.

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. Memory & Cognition, 28(8), 1310-1320.

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. Memory & Cognition, 31(8), 1284-1292.

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.

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

Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activitiy sequences?. Memory & Cognition, 28(2), 214-223.

Sgouramani, H., & Vatakis, A. (2014). “Flash” dance: How speed modulates perceived duration in dancers and non-dancers. Acta psychologica, 147, 17-24.

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

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts.Journal of Memory and Language, 53(1), 125-140.

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. Psychological science, 20(8), 989-999.

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. Cognitive, Affective, & Behavioral Neuroscience, 3(4), 335-345.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. Psychological science, 12(2), 153-156.

Stefanacci, L., Buffalo, E. A., Schmolck, H., & Squire, L. R. (2000). Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient EP. The Journal of Neuroscience, 20(18), 7024-7036.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272.

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. Nature Neuroscience, 16(2), 139–145. https://doi.org/10.1038/nn.3303

Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. Journal of experimental Psychology, 103(2), 255.

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

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep Spindle Activity is Associated with the Integration of New Memories and Existing Knowledge. Journal of Neuroscience, 30(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Taylor, B. (1977). Tense and continuity. Linguistics and philosophy, 1(2), 199-220.

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

Tulving, E. (1985). Elements of episodic memory.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. Psychological review, 94(1), 3.

Van Dijk, T. A., Kintsch, W., & Van Dijk, T. A. (1983). Strategies of discourse comprehension. New York: Academic Press.

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.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. Science, 277(5324), 376-380.

Vega, M. (1996). The representation of changing emotions in reading comprehension. Cognition & Emotion, 10(3), 303-322.

Vendler, Z. (1967). Linguistics in philosophy.

Verkuyl, H. J. (1996). A theory of aspectuality: The interaction between temporal and atemporal structure (Vol. 64). Cambridge University Press.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. Cognitive Psychology, 48(4), 422-488.

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annu. Rev. Psychol.*, *57*, 139-166.

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., & Gennari, S. P. (2019). How language and event recall can shape memory for time. Cognitive Psychology, 108, 1–21. https://doi.org/S0010028518301610

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

Wearden, J. H. (1991). Do humans possess an internal clock with scalar timing properties?. Learning and Motivation, 22(1), 59-83.

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why"sounds are judged longer than lights": Application of a model of the internal clock in humans. The Quarterly Journal of Experimental Psychology: Section B, 51(2), 97-120.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. Psychological bulletin, 121(3), 331.

Wilder, D. A. (1978a). Effect of predictability on units of perception and attribution. Personality and Social Psychology Bulletin, 4(2), 281-284.

Wilder, D. A. (1978b). Predictability of behaviors, goals, and unit of perception. Personality and Social Psychology Bulletin, 4(4), 604-607.

Williamson, R. A., & Markman, E. M. (2006). Precision of imitation as a function of preschoolers' understanding of the goal of the demonstration. Developmental psychology, 42(4), 723.

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

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition, 69(1), 1-34.

Woodward, A. L., & Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. Psychological Science, 11(1), 73-77.

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

Yasuda, K., Watanabe, O., & Ono, Y. (1997). Dissociation between semantic and autobiographic memory: A case report. Cortex, 33(4), 623-638.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. Cognitive Science, 28(6), 979-1008.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. Current directions in psychological science, 16(2), 80-84.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. Psychological bulletin, 127(1), 3.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., ... & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. Nature neuroscience, 4(6), 651-655.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. Journal of Experimental Psychology: General, 138(2), 307.

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.

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. Psychology and aging, 21(3), 466.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. Journal of Experimental Psychology: General, 130(1), 29.

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., & Block, R. A. (1997). Temporal cognition. Current Directions in Psychological Science, 6(1), 12-16.

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.

Zelkind, I. (1973). Factors in time estimation and a case for the internal clock. The Journal of general psychology, 88(2), 295-301.

Zwaan, R. A. (1996). Processing narrative time shifts. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(5), 1196.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. Psychology of learning and motivation, 44, 35-62.

Zwaan, R. A. (2008). Time in language, situation models, and mental simulations. Language Learning, 58(s1), 13-26.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. PloS one, 7(12), e51382

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. Psychological bulletin, 123(2), 162.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. Psychological science, 292-297.

Zwaan, R. A., Madden, C. J., & Whitten, S. N. (2000). The presence of an event in the narrated situation affects its availability to the comprehender. Memory & Cognition, 28(6), 1022-1028.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. Journal of experimental psychology: Learning, memory, and cognition, 21(2), 386.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. Scientific studies of reading, 2(3), 199-220.