

Remembering time: The role of event structure in duration representation

Myrthe Faber

PhD

University of York

Psychology

May 2015

Abstract

This thesis investigates the role of event structure in duration representation. A combination of behavioural and neuroimaging techniques was used to examine the effect of the number of perceived segments and the relative similarity between them on memory representations and estimates of duration. Behavioural studies in Chapter 2 showed that an increase in the number of perceived segments and a decrease in perceived similarity between them lead to longer estimates of duration when reconstructing duration based on a memory representation of content. Chapter 3 investigated whether representations of duration arising from language are similar to those from visual stimuli, indicating that for language, an increase in the number of segments but not the similarity between them leads to longer estimates. Chapter 4 investigated whether event structure also affects time monitoring, showing that estimated duration increases as an effect of the number of segments and dissimilarity between them when both time and content are attended to, but that only the number of coarse segments plays a role when only time is attended to. Together, these findings corroborate the idea that duration reconstruction relies on the encoded event structure, as the role of event structure is diminished when content is not remembered. However, on a coarse level, the number of event boundaries may also guide the encoding of duration. Chapter 5 investigated the neural underpinnings of duration reconstruction using fMRI, showing that activity in left hippocampus is modulated by event structure. Finally, a behavioural experiment in Chapter 6 investigated the effect of event structure on the mental reproduction of events, showing that the duration of this replay increases as an effect of more segments and less similarity between them. Together, these findings suggest that event structure affects memory representations, with more segments and less similarity between them leading to longer duration reconstructions.

Table of Contents

Abstract.....	III
List of Figures	X
List of Tables.....	XI
Acknowledgements.....	XII
Author's Declaration.....	XIII
Chapter 1	1
Remembering and reconstructing duration.....	1
1.1 Time.....	2
1.2 How we experience, remember and reconstruct time.....	3
1.2.1 Studying psychological time.....	3
1.2.2 Early psychophysics of time perception: duration is relative	6
1.2.3 Models of human timing	10
1.2.3.1 Clock-based models of duration perception.....	11
1.2.3.2 Cognitive models of duration perception	14
1.2.3.2.1 Storage Size and Processing Effort models	14
1.2.3.2.2 Change-based and Contextual Change models	16
1.2.4 Prospective and retrospective timing.....	19
1.2.5 The notion of events in time research	23
1.3 Memory and Events.....	26
1.3.1 Episodic memory	26
1.3.2 Link between episodic memory and event models.....	28
1.3.3 Event structure	29
1.3.4 Event segmentation and memory formation	33
1.3.5 Event segmentation and time.....	36
1.4 Putting together the pieces of the puzzle: duration, memory and event segmentation	38
Chapter 2	40
The effect of event structure on reconstructed duration	40
2.1 Introduction	40
2.1.1 Causal structure and event perception	41
2.1.2 Number of event boundaries	42
2.1.3 Similarity between sub-events.....	42
2.1.4 Similarity and time	43
2.2 Experiment 1: Reconstructing duration from memory	45
2.2.1 Research question and aims.....	45
2.2.2 Novel retrospective paradigm.....	46
2.2.3 Stimulus choice	48

2.2.4 Research hypotheses.....	49
2.2.5 Stimulus pre-tests.....	50
2.3 Methods	50
2.3.1 Participants.....	50
2.3.2 Materials	51
2.3.3 Stimulus ratings.....	51
2.3.4 Design and procedure.....	53
2.4 Results	56
2.4.1 Recognition memory	56
2.4.2 Duration ratings	57
2.4.3 Regression results	58
2.5 Discussion	59
2.5.1 Implications of experimental design	60
2.6 Experiment 2: Reconstructing duration from memory: extension of the paradigm.....	61
2.6.1 Research hypotheses and aims.....	61
2.7 Methods	62
2.7.1 Participants.....	62
2.7.2 Materials	63
2.7.3 Stimulus ratings.....	63
2.7.4 Design and procedure.....	66
2.8 Results	67
2.8.1 Recognition memory	67
2.8.2 Duration ratings	68
2.8.3 Regression results	69
2.9 Discussion	70
2.9.1 General discussion.....	70
2.10 Conclusion.....	73
Chapter 3	74
The effect of event structure on reconstructed duration based on linguistic descriptions of events.....	74
3.1 Introduction	74
3.2 Memory for duration and language.....	74
3.2.1 Retrospective interference through linguistic post-event information...	75
3.2.2 Retrospective interference through verbal description	76
3.2.3 Duration estimation based on verbal description	79
3.2.4 Research aims and questions	80
3.3 Experiment 3: Exploration of the effect of event structure on naturalistic descriptions of events.	81
3.3.1 Research hypotheses and aims	81
3.4 Methods	82
3.4.1 Participants.....	82
3.4.2 Materials	82

3.4.3 Design and procedure	82
3.5 Results	83
3.5.1 Number of words	83
3.5.2 Number of verbs	84
3.5.3 Number of different verbs	85
3.5.4 Correlation analyses	86
3.6 Discussion	87
3.7 Experiment 4: Duration ratings based on verbal descriptions	89
3.7.1 Research hypotheses and aims	89
3.8 Methods	90
3.8.1 Participants	90
3.8.2 Materials	90
3.8.2.1 Stimulus creation	91
3.8.2.2 Stimulus pre-test	92
3.8.3 Design and procedure	92
3.9 Results	94
3.9.1 Recognition results	94
3.9.2 Duration ratings	95
3.9.3 Content analyses and regression results	95
3.10 Discussion	97
3.10.1 Stimulus choice	98
3.10.2 Compression of information in vision and language	99
3.11 Conclusion	100
Chapter 4	102
Role of event structure in prospective timing	102
4.1 Introduction	102
4.2 Attending to time	103
4.2.1 Attention and cognitive demands in prospective timing	104
4.2.2 Memory and event complexity in prospective timing	106
4.3 Experiment 5: Remembering event structure in prospective timing	109
4.3.1 Research hypotheses and aims	109
4.4 Methods	110
4.4.1 Participants	110
4.4.2 Materials	111
4.4.3 Design and procedure	111
4.5 Results	112
4.5.1 Recognition memory	112
4.5.2 Duration ratings between conditions	114
4.5.3 Duration ratings between paradigms	114
4.5.4 Regression results	114
4.6 Discussion	116
4.6.1 Comparison between paradigms	116
4.7 Experiment 6: The role of event structure in prospective timing	118

4.7.1 Research hypotheses and aims	118
4.8 Methods	120
4.8.1 Participants.....	120
4.8.2 Materials	120
4.8.3 Design and procedure.....	121
4.9 Results	121
4.9.1 Recognition memory	121
4.9.2 Duration ratings between conditions	123
4.9.3 Duration ratings between experiments	124
4.9.4 Regression results	126
4.10 Discussion	127
4.10.1 Event boundaries as temporal markers	128
4.10.2 Recognition memory and response latencies	129
4.10.3 Working memory versus episodic memory	130
4.10.4 The effect of a secondary task.....	132
4.11 General discussion.....	132
4.11.1 The role of memory in prospective paradigms	132
4.11.2 The role of event structure in prospective paradigms.....	133
4.11.3 The effect of a secondary task in prospective paradigms.....	134
4.12 Conclusion.....	135
Chapter 5	136
The effect of event structure on reconstructed duration: a neuroimaging study..	136
5.1 Introduction	136
5.2 Hippocampus in episodic memory and time	137
5.2.1 Medial temporal lobe and its role in episodic memory.....	137
5.2.1.1 Medial temporal lobe and its role in recollection	137
5.2.1.2 The role of hippocampus in thinking about events.....	139
5.2.1.3 The role of hippocampus in temporal coding.....	142
5.2.1.4 The role of hippocampus in thinking about temporal aspects of past events	147
5.3 Experiment 7: The role of hippocampus in retrospective duration reconstruction.....	149
5.3.1 Research question and aims.....	149
5.3.2 Research hypotheses.....	151
5.4 Methods	152
5.4.1 Participants.....	152
5.4.2 Materials	153
5.4.3 Task design	153
5.4.4 Acquisition parameters.....	154
5.4.5 Data analysis	154
5.4.5.1 Region of interest analysis	156
5.4.5.2 Second-level modulation of event properties.....	156
5.5 Behavioural results.....	157

5.5.1 Recognition memory	157
5.5.2 Duration ratings	159
5.5.3 Regression results	160
5.6 fMRI results.....	161
5.6.1 Recognition memory	161
5.6.1.1 Whole brain results recognition memory	161
5.6.1.2 Region of interest results recognition memory	163
5.6.2 Duration reconstruction	164
5.6.2.1 Whole brain results duration reconstruction.....	164
5.6.2.2 Region of interest results duration reconstruction.....	167
5.6.2.3 Whole brain results duration reconstruction and event properties ..	170
5.7 Discussion	174
5.7.1 Hippocampus and duration estimation.....	175
5.7.2 Event memory circuit and the role of hippocampus, lingual gyrus and amygdala.....	176
5.7.3 Event perception in the brain.....	180
5.7.4 Limitations and further research.....	182
5.8 Conclusion	184
Chapter 6	186
The effect of event structure on reconstructed duration: a mental reproduction study	186
6.1 Introduction.....	186
6.2 Mental replay of events.....	186
6.3 Experiment 8: Reconstructing the unfolding of events from memory.....	187
6.3.1 Research hypotheses and aims	187
6.4 Methods	188
6.4.1 Participants.....	188
6.4.2 Materials	189
6.4.3 Design and procedure.....	189
6.5 Results	190
6.5.1 Recognition memory	190
6.5.2 Mental replay.....	191
6.5.3 Regression results	192
6.6 Discussion	194
6.6.1 Implications.....	194
6.6.2 Further research.....	195
6.7 Conclusion	197
Chapter 7	198
Thesis summary, discussion and conclusions	198
7.1 Summary	198
7.2 Aims and implications	201

7.2.1 General aims.....	201
7.2.2 Implications for theories of time perception	202
7.2.3 Implications for the statistical learning literature	205
7.2.4 Implications for the event memory literature.....	206
7.2.5 Implications for the semantic memory literature	208
7.2.6 Implications for the hippocampus literature.....	210
7.3 Limitations and further research	211
7.4 Conclusions	214
Appendix A.....	216
References.....	219

List of Figures

Figure 1. Information processing model of Gibbon & Church (1984).	12
Figure 2. Diagram depicting Event Segmentation Theory.	33
Figure 3. Example of animation triads.....	46
Figure 4. Schematic overview of the experiment.....	55
Figure 5. Response latencies over subjects.	57
Figure 6. Duration ratings over subjects (scale 1-7)..	58
Figure 7. Example of animation triads.....	65
Figure 8. Response latencies over subjects..	67
Figure 9. Duration ratings over subjects (scale 1-7)..	68
Figure 10. Average number of words per item per condition.....	84
Figure 11. Average number of verbs per item per condition.	85
Figure 12. Average number of different verbs per item per condition.	86
Figure 13. Response latencies over subjects.....	94
Figure 14. Duration ratings over subjects (scale 1-7).....	95
Figure 15. Response latencies over subjects.....	112
Figure 16. Duration ratings over subjects (scale 1-7).....	113
Figure 17. Response latencies over subjects.....	122
Figure 18. Duration ratings over subjects (scale 1-7).....	123
Figure 19. Ratios between estimated duration and actual duration over items.....	124
Figure 20. Duration ratings over subjects (scale 1-7).....	125
Figure 21. Response latencies over items.	158
Figure 22. Duration ratings over subjects (scale 1-7).....	160
Figure 23. Whole brain results for probe recognition over baseline	162
Figure 24. Whole brain results for duration reconstruction over baseline.....	165
Figure 25. Contrast basic < dissimilar in left hippocampus for duration reconstruction.	167
Figure 26. Spherical ROI (9mm radius) drawn around the local maximum within the left hippocampus.....	168
Figure 27. BOLD percent signal change extracted from the spherical ROI.....	168
Figure 28. Contrast basic < dissimilar for duration reconstruction on a liberal threshold.....	169
Figure 29. Whole brain results for number of segments and perceived similarity .	172
Figure 30. Whole brain results for an increase in the number of segments and a decrease in similarity.....	174
Figure 31. Response latencies over items.	191
Figure 32. Proportional duration reproductions.....	192

List of Tables

Table 1. Mean ratings of number of event boundaries and similarity.....	53
Table 2. Regression coefficients for duration ratings..	59
Table 3. Mean ratings of number of event boundaries and similarity.....	66
Table 4. Regression coefficients for duration ratings.	70
Table 5. Regression coefficients for duration ratings.	97
Table 6. Regression coefficients for duration ratings.	115
Table 7. Regression coefficients for duration ratings.	127
Table 8. Regression coefficients for duration ratings.	161
Table 9. Recognition memory over baseline.....	162
Table 10. Duration reconstruction over baseline.	166
Table 11. Whole brain results for number of segments.....	171
Table 12. Whole brain results for perceived similarity.....	171
Table 13. Whole brain results for an increase in the number of segments and a decrease in similarity	173
Table 14. Regression coefficients for duration reproductions.....	193

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr Silvia Gennari, for her continuous support throughout my Ph.D. journey. Serendipitous circumstances have brought us together during an extended study visit that was part of my Master's degree, and I am very grateful that she then accepted me as a Ph.D. student. Her endless enthusiasm and her care for both her research and her Ph.D. students have made it a great journey together. I could not have wished for a better supervisor.

Furthermore, I would like to thank Dr Shirley-Ann Rueschemeyer and Dr Tom Hartley for their valuable advice and guidance as members of my Thesis Advisory Panel. I would also like to thank other members of the lab group and the Department of Psychology and the York Neuroimaging Centre for their advice, comments and suggestions.

My thanks also go out to all of the students who helped with the data collection, and to the over 700 participants who took part in the studies and pre-tests presented here. My special thanks go out to Ada Cirlia and Silvia Harvey, who both volunteered as research assistants.

Finally, I would like to thank my friends and family for their never-ending support. My parents, for always believing in me, being there for me, and teaching me to be curious and critical. My late grandparents, for encouraging and enabling me to pursue my dreams. Pleuni, for being like a sister to me. And my fiancé Koen; I could not have wished for a better person by my side during this and many more journeys.

Author's Declaration

This thesis comprises the candidate's own original work and has not, whether in the same or different form, been submitted to this or any other University for a degree. The work was generously supported by a departmental studentship from the Department of Psychology at the University of York. In addition, I am grateful to the Experimental Psychology Society for awarding Grindley Grants to attend their EPS Meeting in Kent and the Annual Meeting of the Organization for Human Brain Mapping, and to the British Association for Cognitive Neuroscience for awarding me with a studentship to attend their annual meeting. Selected aspects of the research described in this thesis have been published and presented at conferences.

Journal papers:

Faber, M. & S.P. Gennari (2015) In search of lost time: reconstructing the unfolding of events from memory. *Cognition*, 143, p. 193-202.

Faber, M. & S.P. Gennari (2015) Representing time in language and memory: the role of similarity structure. *Acta Psychologica*, 156, p. 156-161.

Conference talks:

Faber, M. & S.P. Gennari (2014) The role of event complexity in memory for event duration. *Experimental Psychology Society*, Kent, UK.

Conference posters:

Faber, M. & S.P. Gennari (2014) The role of event structure in memory for event duration. *British Association for Cognitive Neuroscience Annual meeting*. York, UK.

Faber, M. & S.P. Gennari (2014) Activity in the hippocampus is modulated by event complexity when reconstructing duration. *Organization for Human Brain Mapping Annual Meeting*, Hamburg, Germany.

Faber, M. & S.P. Gennari (2013) The role of event complexity in memory for event duration. *Donders Discussions*, Nijmegen, The Netherlands.

Faber, M. & S.P. Gennari (2012) Retrieving Perceived Duration. *White Rose Doctoral Training Centre Spring Conference*, York, UK.

Chapter 1

Remembering and reconstructing duration

In many situations in everyday life, our memory of how long things take is crucial. For example, in witness testimony, verdicts can critically rely on the memory of the witness of how long something took, like for example how long an encounter with the suspect was. However, we do not always have a measure of duration available in terms of clock time. Instead, we then rely on a sense of how long we *think* things took. Intuitively, we seem to reconstruct duration based on our memory of the events that unfolded during the interval that we are trying to estimate. But what cognitive and neuronal mechanisms underlie this ability? Are these duration estimates veridical, or are they subjective, and affected by the content of the interval? Is there a difference between cognitively timing an interval and reconstructing it from memory? And when we describe what happened to someone else, does what we tell about the content affect how long they think the interval was? Bringing together literature on time perception, episodic memory, event perception and linguistics, this thesis is concerned with answering these questions through a series of experiments investigating the effect of event structure in duration representation.

This chapter will outline the key theoretical frameworks that are relevant to answering these questions. Firstly, this chapter will address the historical development of psychological research into time perception and observations from early psychophysical studies into time perception, suggesting that time perception is subjective and dependent on characteristics of the stimuli. Secondly, this chapter will give an overview of theories on human timing that have been developed. Both biologically motivated clock-based models and cognitive models will be discussed, focusing on the relevance of the latter for answering the questions above. Thirdly, this chapter will address crucial distinctions between prospective and retrospective research methodologies, arguing that retrospective paradigms can shed light on the memory mechanisms that play a role in reconstructing duration.

Furthermore, this chapter will discuss episodic memory and event segmentation, linking the two by highlighting how event segmentation may determine episodic memory content. Finally, this chapter will outline how event segmentation may affect duration reconstruction, and how this thesis will further investigate this possibility.

1.1 Time

The concept of time has captured the imagination of many scientists, philosophers and artists. From philosophical views on time as an absolute such as the Kantian view that time is an inner sense that has empirical existence (1770) to views that express the relativity of time in science and physics such as Albert Einstein's famous quote that "time is relative" (1920) to artistic interpretations like Salvador Dalí's painting *The Persistence of Memory* (1931) that shows a melting pocket watch signifying the "softness" of time, views on time and its nature have evolved and changed over time. Questions about time and its nature are as old as mankind, starting with early civilizations studying the passage of the seasons. The first philosophical ideas about the nature of time came from the early Greek philosophers. Among these was Heraclitus (c. 535-475 B.C.) who believed that permanence did not exist and that only change had reality (in: Roeckelein, 2008). He reasoned that the only real state is the state of "becoming". The philosopher Plato (427-347 B.C.), in contrast, believed that time exists, and that it flows independently from the events occurring in time (in: Roeckelein, 2008). The first philosopher to ask how we perceive time was Aristotle (384-322 B.C.). He hypothesised that time is dependent on aspects of motion, such as "before" and "after" (in: Roeckelein, 2008). He reasoned that, therefore, our perception of time is dependent on the "number of motion" (in: Roeckelein, 2008). As the following sections will illustrate, these ideas are still, in a slightly altered form, at the heart of the debate on how we perceive time.

1.2 How we experience, remember and reconstruct time

There are many ways to keep track of time, such as counting seconds or using a timing device such as a watch. But how do we build a sense of time when we did not count or have access to a timing device, or even when we did not attend to time at all in the first place? This is known as *psychological time* and is defined as a subjective estimate of time, without the use of an external timing device or environmental clues such as the position of the sun (English & English, 1958, in: Roeckelein, 2008). As the following sections will outline, there is a lot of discussion about how exactly we experience time from a psychological point of view. The following sections will provide an overview of the development of different theories on psychological time.

1.2.1 Studying psychological time

The concept of psychological time was born at the end of the 19th century (Roeckelein, 2008). In 1890, William James (1842-1910) introduced the notion of psychological time in his book *The Principles of Psychology*. James's main postulation was that in order to understand the past, one must relate this past to the present, and one must understand what the present is. In his book, he therefore discussed three concepts related to the notion of the psychological experience of time. The first concept is the "specious present", which refers to our experience of the "now". When we focus on the "now", the present seems to expand. Secondly, James discussed the concept of "successiveness", claiming that only through a feeling of succession one can experience a feeling of past. Thirdly, James discussed the perception of time in terms of retrospection (versus prospection). The past is by definition retrospective, and as such, "the perception [of past time] goes by the name of *memory*" (James, 1890/2007, p. 605). When we are paying attention to time, the duration of a period of time is perceived differently from the same period of time in retrospect when we did not pay attention to time. What happened in this period of time determines whether we perceive it as longer or shorter:

In general, a time filled with varied and interesting experiences seems short in passing, but long as we look back. On the other hand, a tract of time empty of experiences seems long in passing, but in retrospect short (James, 1890/2007, p. 624).

Furthermore, James concluded that awareness of change is crucial for our experience of time, and that these changes must be of some concrete sort. In sum, by focusing on the characteristics of the stimuli such as interestingness and (the number of) changes, James was among the first psychologists to conclude that there must be a relationship between our experience of duration and the filling of the duration.

Around the same time, a book on the concept of time from a psychological perspective by the French philosopher Jean-Marie Guyau (1854-1888) was published. His approach was ahead of its time, as he thought about the perception of time in relation to human information processing, and how concepts of time develop from childhood to adulthood (Ornstein, 1969; Block & Zakay, 2001). Guyau considered time to be a purely mental construction from the events that take place (1890, as paraphrased by Ornstein, 1969, p. 37). In other words, he postulated that time itself does not exist, but that it only exists by the events that occur in it, and that time is a product of human imagination, memory and will (Ornstein 1969, Roeckelein 2008). According to Guyau, our experience of time is based on characteristics of the stimuli, such as the intensity of the stimuli, the extent of the differences between the stimuli, the number of stimuli, the attention paid to the stimuli, the associations of the stimuli and the expectations called up by the stimuli (Ornstein 1969; Roeckelein 2008). He hypothesised that these factors can expand our perceived duration, and stated that it is an increase in the number of “mental images” that lengthens our perceived duration (Ornstein 1969, p. 38).

Note that both James and Guyau hypothesised that our perceived or experienced duration is dependent on the characteristics of the stimuli (i.e. what happened in an interval of time). However their ideas were mainly philosophical, as neither of them did any empirical research. Their contemporary Wilhelm Wundt

(1832-1920) was among the first to study the experience of time experimentally. Wundt and colleagues investigated properties of the filled duration illusion (where a filled duration seems longer or shorter than an empty or unfilled duration) and determined minimum and maximum durations and intervals for this illusion (Roediger 2008). In his book *Outlines of Psychology* (1897), Wundt noted that duration must therefore be a relative concept (Wundt, 1897/2009, p. 144). Each moment in time is filled by a certain content that has a relationship to the subject experiencing the flow of time. Even when the sensational content remains identical over time (e.g. when a person experiences a state, or a lasting impression), the relationship to the subject cannot be constant as his or her internal states are ever changing. Therefore, Wundt was the first to argue that the experience of duration cannot be absolute and must be subjective, based on empirical evidence.

An important step forward in research into psychological time was the publication of the first review of time perception literature in 1891. Nichols (1852-1936) was the first scholar to not only review the existing literature on time perception to date, but also the first to publish this in the only psychological journal at that time, the *American Journal of Psychology* (Block & Zakay, 2001). Because of its publication in this medium, the paper received widespread attention from researchers in the field of psychology. In his review, Nichols gave an overview of the historical development of (philosophical) theories on time perception, and summed up the experimental research to date. As a conclusion of his historical overview, Nichols wrote the following:

Casting an eye backward we can but be struck by the wide variety of explanations offered for the time-mystery. Time has been called an act of mind, of reason, of perception, of intuition, of sense, of memory, of will, of all possible compounds and compositions to be made up of them. It has been deemed a General Sense accompanying all mental content in a manner similar to that conceived of pain and pleasure. It has been assigned a separate, special, disparate sense, to nigh a dozen kinds of 'feeling', some familiar, some strangely invented for the difficulty. It has been explained by

'relations', by 'earmarks', by 'signs', by 'remnants', by 'struggles', and by 'strifes', by 'luminous trains', by 'blocks of specious-present', by 'apperception'. It has been declared *a priori*, innate, intuitive, empirical, mechanical. It has been deduced from within and without, from heaven, and from earth, and from several things difficult to imagine as of either (Nichols, 1891, p. 502).

This conclusion sums up many of the issues that are still relevant to the current debate on time perception. As illustrated above, James, Guyau and Wundt stood at the beginning of the tradition of studying of time from a psychological point of view. They shared the view that our experience of duration is relative and that it depends on some sort of characteristic of the filling of the duration. As will be discussed in the remainder of this chapter, these notions are still at the centre of the current debate on how we perceive, experience, remember and estimate time. The following section will give an overview of the experimental research that tried to determine these characteristics using early psychophysical techniques.

1.2.2 Early psychophysics of time perception: duration is relative

Psychophysics studies the relationship between (characteristics of) a stimulus and behaviour or, in this, case perception. Many researchers have attempted to determine the nature of the relationship between real time and our perception of time. Early time research often explained this relationship using Vierordt's law (1868), that states that short intervals are overestimated and long intervals are underestimated (Woodrow, 1951, in: Allan, 1979). Furthermore Vierordt's law predicts that there is a point or interval of indifference at which people do not overestimate or underestimate an interval (Woodrow, 1951, in: Allan, 1979). According to Vierordt, at the point of indifference the experienced time equals the physical time, meaning that the estimated time signifies an actual time percept (in: Eisler, Eisler, & Hellström, 2008). Through a series of experiments using auditory stimuli, Woodrow (1930, 1933, 1934) showed that in time perception, there is no

point of indifference. He argued that if a point of indifference is found, it varies too much from other points of indifference found in other laboratories under the same experimental conditions, and that there is no point of indifference that is valid for all experimental participants (reviewed in: Allan, 1979). He therefore concluded that Vierordt's law is not applicable to our experience of time, implying that there is no point of indifference where our subjective reproduction of time equals our direct time percept.

The question then arises whether the point of indifference exists at all, and if so, why varying points of indifference are reported between studies. One of the factors that has been shown to affect the point of indifference is the range of stimuli used in an experiment. This was shown by Fraisse (1948), who asked people to reproduce the durations of streams of sounds. He showed that the point of indifference depends on the range of stimuli used: for instance, for a scale of 0.2 to 1.5 seconds the point of indifference is different from the indifference point for a scale of 0.3 to 12 seconds (Fraisse, 1948, in: Eisler et al., 2008). These findings suggested that any range of stimuli could generate a point of indifference, and that any point of indifference must be relative to the scale it is on, challenging the idea that a point of indifference equals a direct percept of time. In other words, there is no absolute duration at which our perception of duration always equals the actual duration.

In addition, there are numerous other principles that appear to affect the relationship between actual duration and our percept of it. For instance, when a participant is exposed to a series of stimuli on a scale, their memory of these stimuli is likely to be shifted towards the middle of the scale. This idea was formulated by Leuba (1893) and is known as the *law of sense memory* (in: Eisler et al., 2008). The law of sense memory furthermore states that our memory of stimuli shifts toward what has been experienced most frequently (Leuba, 1893, in: Eisler et al., 2008). In contrast, Hollingworth (1909, 1910 in: Eisler et al., 2008) observed that it is not our memory of the stimuli that shifts towards a centre, but that it is rather our judgement that is subject to a central tendency. Whether it is our memory of the stimuli or our judgement, either way the behavioural outcome

displays a bias to the mean, again suggesting that points of indifference are relative rather than absolute (Eisler et al., 2008).

Another important observation was made by Fechner (1860). Studying people's judgement of whether a weight is heavier than a standard weight, he found that it makes a difference whether one compares a stimulus to the previous stimulus, or to the next stimulus. This phenomenon is also known as the *time-order error*. The time-order error has been shown for almost every modality, including duration perception (an overview of studies investigating time-order error for duration can be found in: Eisler et al., 2008). The time-order error primarily occurs in tasks that use a forced-choice paradigm. In a forced-choice task, there are normally two orders in which the stimuli can be presented: the shorter of the stimuli is presented first and then the longer, or the longer of the stimuli is presented first and then the shorter. Early studies found that for short durations, the time-order error is often positive, while for longer durations it is often negative (summarised in: Allan, 1979). This indicates that there is more to subjective duration than Vierordt's law, namely that our judgement of duration is affected by the order in which and the scale on which stimuli are presented. However, the nature of the time-order error has been under debate: on the one hand there are researchers who support the idea that the time-order error is a perceptual phenomenon that arises from adaptation to the set of stimuli presented and the magnitude estimations associated with them (e.g. Hellström, 1977), while other researchers claim that the time-order error is an artefact of the forced-choice task and thus depends critically on the response type used in a paradigm (e.g. Allan, 1977, both in: Fraisse, 1984, p. 11).

As noted above, psychophysical research has shown that duration sensation and perception are subject to expansion and contraction dependent on the scale, task and clock duration. Further psychophysical research has focused on exactly how a participant's behavioural response is related to the intensity of a stimulus. One of the most well known psychophysical laws that describes this relationship is Weber's Law (1834). Weber's Law states that the threshold of discrimination between two stimuli, such as brightness, loudness, or duration, increases

monotonically as the intensity of stimuli increases (as described by: Grondin, 2001). Attempting to define the exponent that explains the relationship between subjective duration and actual duration (i.e. a constant proportion that defines the ratio between subjective and actual duration), Eisler (1976) conducted a meta-analysis of studies of subjective duration published between 1868 and 1975, and found that an average exponent of .9 seems to best describe the ratio between subjective duration and actual duration. He claimed that the “comparatively small variation in the exponent indicates an astonishing stability in the data for time perception” (Eisler, 1976, p. 1157). However, his meta-analysis did not distinguish between different methods of measurement and different stimulus conditions such as the exact length of the durations used in the experiment, which, as argued above, can greatly influence the results found in a particular experiment.

In an attempt to replicate the findings reported by Eisler (1976), Allan (1978) found that different values are obtained between subjects and between tasks. In her experiment, participants were asked to reproduce ten intervals, to reproduce a duration of half these intervals and to reproduce a duration of double these intervals. These data showed that the exponents estimated from twelve subjects varied from .50 to 1.30, and that the correlation between the reproduction exponents between the tasks was low. Together, the findings described above illustrate that although there is no widespread agreement on the exact nature of the relationship between the intensity of a stimulus and human sensation, the results from psychophysical studies on duration experience generally suggest that regardless of the exact experimental conditions, there is a robust and persistent discrepancy between stimulus duration and subjective duration.

Moreover, Allan (1979) pointed out the following: in all of the published studies of time reviewed in her 1979 study that report verbal and scalar (magnitude) estimates of clock duration, there has been the *assumption* that there is an empirical relationship between the estimates given by a participant and the actual stimulus duration that directly reflects the psychophysical law. Implicitly or explicitly, studies have assumed that this obtained mathematical function represents the transformation of stimulus time to perceived time, and that

therefore a response given by a participant is a simple (linear or otherwise) transformation of perceived time (Allan, 1979, p. 343). However, this relationship between real time and experienced time may in general not be as straightforward as the early psychophysical studies assumed. Terms like perceived, internal, subjective, psychological, and apparent duration or time are used as synonyms referring to a temporal value provided by a person to make a temporal judgement (summarised in: Allan, 1979). However this does not directly imply that this temporal value is a representation of 'real' time experience, or, as Efron (1970) argued: that "the duration of a percept is not necessarily the same as perceived duration" (Efron, 1970, as cited by: Allan, 1979, p. 341).

Similarly, Ornstein (1969) made the observation that early psychophysical research was mainly focussed on how accurately 'real' time is perceived. He argued that "calling the clock of hours, minutes and seconds 'real' time is like calling American money 'real money'. An analysis should be concerned with experiential time *per se*, not as it might relate to hours, days, burning rope or to some other time definition" (Ornstein, 1969, p. 20). In sum, the relationship between stimulus time and 'perceived' time may not be as direct as early models assume. Perceived duration appears to be subjective, relative and dependent on stimulus characteristics. The question that remains, however, is how humans build this experiential sense of time and what stimulus properties affect duration estimation.

1.2.3 Models of human timing

In order to address how humans and animals build a sense of time, many models of time perception have been proposed. Block and Zakay (1997) pointed out that it is puzzling that there is no single sensory organ or perceptual system that is responsible for our psychological experience of time, and observed that therefore research has focussed on the interplay between cognition and biological processes that contribute to human timing. Models of human timing can roughly be divided into (at least) two categories: clock-based models and cognitive models. The following sections will give an overview of the development of clock-based models

and cognitive models and will highlight some of the most prevalent models of human timing.

1.2.3.1 Clock-based models of duration perception

The idea of a clock-based model was introduced in the 1960s. Classical clock-based models assume that humans have a pacemaker-accumulator system. A pacemaker emits 'ticks' that are collected by the accumulator that counts the number of ticks. The more ticks are collected, the longer the duration (Grondin, 2001). Two of the earliest models were formulated by Creelman (1962) and Treisman (1963). Both models propose a pacemaker-accumulator system, but have different assumptions about whether and how external perturbations affect the frequency with which the pacemaker emits pulses. Creelman's model assumes that the pacemaker emits pulses with a fixed frequency, which means that one can predict how many pulses there are per unit of time (Grondin, 2001). In contrast, Treisman (1963) proposed a model in which a pacemaker emits pulses at a variable frequency that is subject to external factors (Grondin, 2001). So even though both models propose a pacemaker-accumulator model, they differ in whether the frequency at which this pacemaker emits pulses is fixed or subject to external factors causing it to be variable.

An important extension of the internal clock model was proposed by Gibbon, Church and Meck (1984) (Figure 1). Scalar expectancy theory (SET) was originally developed to explain animal learning. Similar to the earlier model proposed by Treisman, SET proposes a continuously running pacemaker that emits pulses at a stochastic rate. This pacemaker is connected to an accumulator by a switch that only allows pulses to flow to the accumulator when a stimulus is timed. When the stimulus timing process ends, the number of pulses stored in the accumulator forms the basis for a duration judgement (as described by: Matthews, Stewart, & Wearden, 2011). The content of the accumulator is stored in working memory, and a reference memory that gives an animal information about whether a certain count value leads to a reward (Grondin, 2001, p. 29). Therefore, a third

mechanism, namely the comparator, is necessary in the SET model, as the count value of an individual stimulus needs to be compared to the count value that leads to a reward (Grondin, 2001, p. 29). SET allows for the frequency of the pacemaker to be variable. For example, the rate of the pacemaker could be influenced by the stimulus intensity (Zelkind, 1973) and switch latencies (Wearden, Edwards, Fakhri, & Percival, 1998, both in: Matthews et al., 2011).

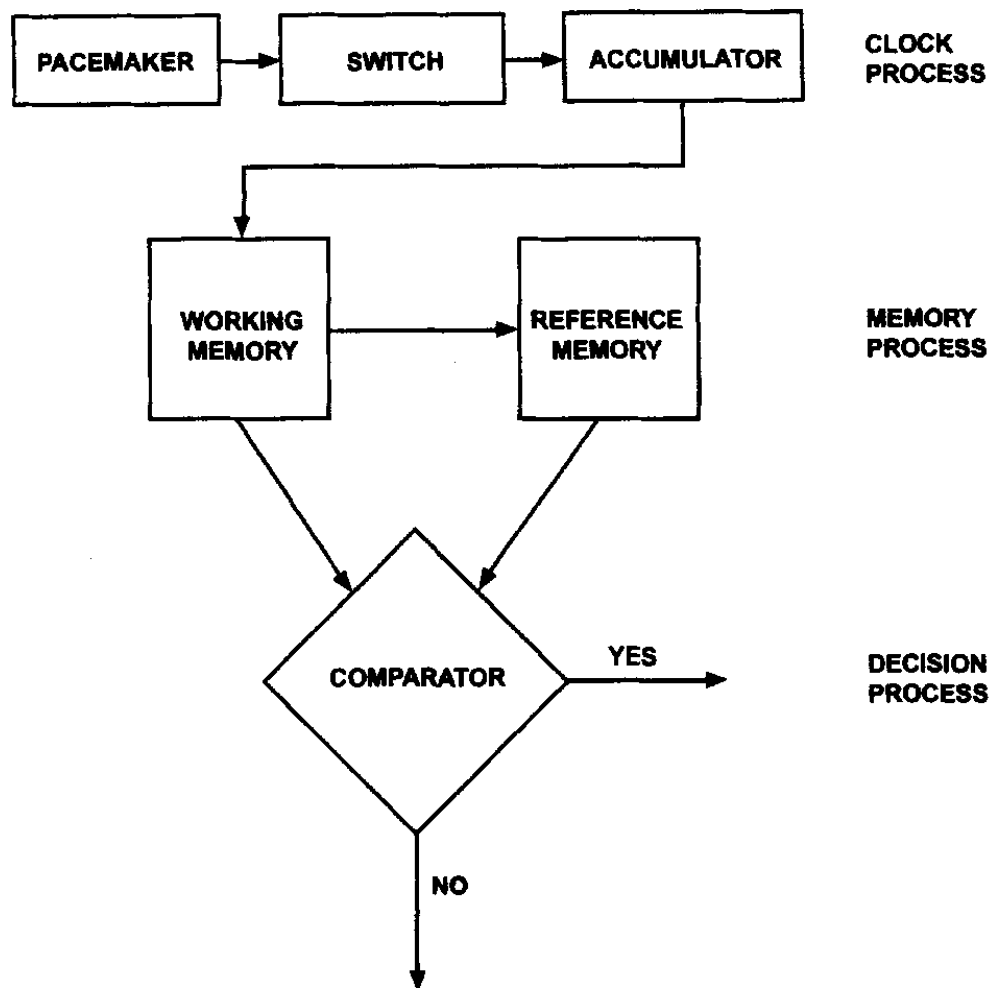


Figure 1. Information processing model of Gibbon & Church (1984). The figure, adapted from Grondin (2001, p. 28), shows three processes: a clock process, a memory process and a decision process. The clock process contains a pacemaker component, a switch and an accumulator. The switch connects the pacemaker with the accumulator, allowing pulses only to flow to the accumulator when a stimulus is timed.

Wearden (1991) was the first to successfully apply this model to human behaviour as well, suggesting that humans have a scalar representation of duration. Wearden used a temporal generalisation paradigm in which participants are familiarised with a standard duration. Participants were then asked to compare new stimuli to this standard duration and tell whether they are the same. Wearden argued that although humans perform differently from animals in these temporal generalisation experiments, this is due to the fact that the way that humans make judgements about intervals differs from the way animals do, and that the underlying representation of duration that humans have is consistent with SET. He argued that human behaviour displays scalar properties (the mean time intervals that people produce are generally well matched with the clock time, and that the standard deviation of this response increases linearly with the mean) and therefore argued that, in some circumstances, human timing can be described by the SET model (Wearden, 1991).

Clock-based models, however, have been argued to have some limitations. Based on a meta-analysis of 110 studies investigating the psychophysics of time perception through a variation of timing tasks in humans and animals, Gibbon and colleagues (1997) claimed that different time ranges appear to require different processes. Especially timing of longer intervals (more than half a minute) is difficult to explain within a clock-based framework. Moreover, their results suggested that intervals of up to 1-2 seconds might be processed by different neurobiological mechanisms than those above. Cognitive, 'non-temporal' processes may add to the increase in variability observed for durations above the 1 or 2 second range.

Furthermore, Gibbon and colleagues (1997) observed that different timing systems appear to have different neural substrates, and argued that this may challenge the idea of one dedicated neurobiological timing system. Further evidence for this idea came from experiments showing that timing can be disrupted by a secondary cognitive task (e.g. an attentionally demanding task), suggesting that timing shares resources with other cognitive mechanisms (for an extensive review, see: Brown, 2008). So although clock-based models may be valid

for (very) short durations, longer durations appear to recruit cognitive mechanisms.

1.2.3.2 Cognitive models of duration perception

Therefore, instead of focusing on a clock-based timing system in humans, some research has focused on time perception from a cognitive perspective. These models often consider memory and attention to explain how we perceive and remember duration. Theories of cognitive timing roughly consider three areas. Storage size models are concerned with how information is organised in memory, suggesting that when a stimulus requires more mental space, it will be judged as being longer. This is closely related to processing effort models that suggest that it is about the familiarity of stimuli and the ease of encoding, rather than their storage space in terms of stimulus quantity per se. Thirdly, change based models explain subjective duration in terms of the amount of change (both change gleaned from the stimulus as well as contextual change) that needs to be remembered. The following gives a brief overview of some of the most influential models.

1.2.3.2.1 Storage Size and Processing Effort models

Arguably the most widespread theory from the Storage Size perspective was formulated by Ornstein in 1969. He hypothesised that perceived duration is directly related to the storage size occupied by events in memory (Ornstein, 1969). Ornstein argued that if more storage space is used to encode a stimulus, the subjective duration is lengthened. This does not mean that “more” necessarily means “longer”: Ornstein argued that not only the actual information that is stored is important, but also how that information was stored. A certain amount of information can be stored in different ways, depending on how the information is chunked by the memory system. To explain this, Ornstein used the following example:

Suppose you are asked to remember the following sequence of four binary digits:

1010 0100 1111 0111 1101

If you have no coding scheme for this sequence then this sequence would subtend twenty metaphorical spaces in storage. One space for each number in the input array. Suppose I give you a code. In this notation the zero always means zero and the ones, reading from right to left represent powers of 2. So the first (right hand) space if it is a 1 is 2^0 or 1, the next 2^1 or 2, the next 2^2 or 4 and the fourth in each sequence 2^3 or 8. So knowing the code you can store each four binary digit sequence this way: 10 4 15 7 13. Now the whole sequence subtends only eight metaphorical spaces and the stored information is the same, since if you know the code you can produce 0100 from the 4 which is in storage.” (Ornstein, 1969, p. 42).

Ornstein argued that this extends to real life, where based on people’s specialities and experiences they can code certain events more efficiently. In other words, Ornstein’s hypothesis stated that the experienced duration of an interval is a construction based on its storage size, which is affected by the efficiency with which information can be encoded.

Through a series of experiments, Ornstein (1969) gathered support for his hypothesis. He demonstrated that an increase in the number of stimuli, for example a series of consecutive tones that are presented in a certain interval, lengthens the experienced duration. Furthermore, Ornstein investigated to what extent the complexity of a stimulus affects the experience of duration using stimuli like abstract line drawings and tones. He found that complexity increases the experience of duration: the higher the complexity, the longer the duration. He established this effect for abstract drawings with more or less lines and angles, sounds that are more or less easily coded (for example the quick turn of a typewriter roller versus two hair brushes brushing together) and art works of different complexities. Based on these results, Ornstein argued that an increase in the number of events causes the necessary storage space to increase and thus the

experienced duration to increase. He also argued that more storage space is necessary to store increasingly complex events or an increasingly complex sequence of events and thus the experienced duration is lengthened.

There is some additional evidence available to support Ornstein's storage size hypothesis. Berg (1979) studied the effect of chunking on the estimated duration using whether information can be encoded following an organisational schema or not. More specifically, he used social scripts as an organisational principle that could reduce the amount of information that needs to be encoded. Participants watched animations of moving geometric shapes (with a duration of .58 to 3.0 seconds) of which they had to estimate the duration. They were either told that the shapes were people doing things, or that the shapes were moving randomly. Berg's results showed that for durations over 1.6 seconds, the socially framed stimuli obtain shorter duration estimates. Berg argued that this is because the labels that are provided by a social framework could help in summarising and organising the information in an efficient way, reducing the storage size necessary to represent the stimulus. Similar evidence comes from a study by Mulligan and Schiffman (1979) who found that when participants are asked to remember a set of complex line drawings that are ambiguous and uninterpretable by themselves or verbal descriptions that do not make sense on their own, participants who were given a "code" to simplify the stimuli or, in the case of verbal descriptions, a relevant context reported shorter durations than participants who were not given one. This is consistent with Ornstein's (1969) observation that when information can be encoded more efficiently, the storage size decreases and the experienced duration is shorter compared the duration attributed to a stimulus that can be encoded less efficiently and therefore requires more storage space.

1.2.3.2.2 Change-based and Contextual Change models

Similarly, Fraisse (1963) argued that subjective duration is proportional to the number of perceived changes (in: Fraisse, 1984). He claimed that these changes could either be "the number of events that take place in the outside world and that

are perceived, or else the number of events that are identified at a perceptive level and then memorized” (Fraisse, 1984, p. 28). In other words, the internal structure of a witnessed event that has been memorised could affect the subjective duration reported by a person. This is very similar to Ornstein’s hypothesis: according to Fraisse, “Ornstein has adopted the same hypothesis, but has better spelled out the nature of the changes” (Fraisse, 1984, p. 20). He added to Ornstein’s hypothesis claiming that although in some cases duration might be constructed based on the “objective number of changes” (cf. Ornstein, 1969), people might base their duration estimates on the “number of changes memorized in an individualized manner” in other (real life) situations (Fraisse, 1984, p. 20).

This notion of individualised manners of memorising was also adopted in Block and Reed’s contextual change model (1978). They called their model a *contextual* change model because “it emphasizes factors surrounding an event or episode which influence an organism’s encoding of, conceiving of, and responding to the event or episode” (Block, 1990, p. 29). Like Fraisse (1963), Block and Reed argued that important changes during an interval affect how long we estimate it to be (in: Block, 1990). Block and Reed found that expansion in duration estimation occurs when changes in “process context” have occurred, for example when a participant has to use different cognitive processes and encoding strategies (Block, 1990, p. 24). They argued that this information is encoded and stored as part of the representation of the event in memory (Block, 1990, p. 24). This includes changes in task, but also for example environmental changes such as changing location. This means that when a person retrieves information on the duration of an interval, the context information is also retrieved. Thus, if more changes in context are encoded, a participant will use this information to base his or her duration estimate on as well. This also means that personal encoding strategies and personal experiences (for example being familiar with surroundings or not) could affect a person’s duration estimates besides the content of the interval and the type of judgement given (Block, 1990). However, as these contextual changes include changes in environmental context, mood and type of

processing, there is no independent way of measuring the amount contextual change, which is an issue for the contextual change model (Block, 1990).

Poynter (1989) introduced a particular version of a change-based model, in which he emphasised the role of what he calls 'events'. He proposed that the content of an interval defines duration estimates, and that this content is defined by the amount and magnitude of "sensory" change that has been experienced. Furthermore, he argued that the organisation of events should affect duration judgements. He argued that in every day life, the "salience of events filling an interval, how discrete they are and how they are organised in the interval are factors which affect the remembered duration of an everyday interval of time" (Poynter, 1989, pp. 314–315). He proposed that duration judgements depend on the number of perceived events, the discreteness of the events and how memorable they are, depending on how easily the pattern of events can be "chunked" or reduced for efficient storage and retrieval (Poynter, 1989, p. 316). However, as will be outlined below, evidence for this model has been limited in scope, not addressing events as identified discrete segments of everyday experience. Rather, evidence has been disparate, addressing only aspects of the model. However, the following will also argue that Poynter's model proves a fruitful starting point for building a more specific model of duration estimation that relies on insights from memory, perception and time estimation literature.

In sum, cognitive models of duration perception are concerned with the way in which information is encoded, stored, and retrieved. All models appear to have in common that more stored information seems to lead to longer duration estimates. However which properties of the stimuli exactly lead to 'more information' or when stimuli can be encoded 'more efficiently' has not been defined clearly and univocally by these models. The following section will therefore review some of the experimental evidence from studies investigating what properties of the stimuli and task affect prospectively experienced and retrospectively reconstructed duration.

1.2.4 Prospective and retrospective timing

As outlined above, cognitive models of duration perception are concerned with the mechanisms underlying duration estimation from memory. These models attempt to describe the relevant storage and retrieval mechanisms that participants use to reconstruct duration based on the information stored in memory. Hence, these duration reconstructions and subsequent judgements are retrospective in nature. As pointed out above, James (1890) was the first to distinguish between a prospective and retrospective sense of time. In experimental practice, this distinction refers to whether a participant is aware of the fact that a temporal interval will need to be judged (prospective paradigm) or whether the participant is naive to this and will be surprised with an estimation task after the interval or intervals. In the prospective paradigm participants are enabled to intentionally encode temporal aspects of the stimuli, or explicitly told to “time” or measure the stimuli. Block (1990) therefore argues that in the prospective paradigm people make an estimate of *experienced duration* (in: Block & Zakay, 1997). In the retrospective paradigm however people retrieve whatever information they find relevant from memory to make a duration judgement (in: Block & Zakay, 1997). Therefore, in the retrospective paradigm people make an estimate of *remembered duration* (Block & Zakay, 1997). This distinction is crucial, as different variables affect the retrospectively remembered and prospectively experienced (judgements of) duration (Block & Zakay, 1997; see below).

Some studies have explicitly compared prospective and retrospective paradigms in order to investigate the similarities and differences in mechanisms underlying both ‘senses’ of duration. One of the most important studies in this field was done by Hicks, Miller and Kinsbourne (1976) who investigated the difference in judgment between prospective and retrospective timing. In their study, participants were asked to sort playing cards (four suits, two colours). Three conditions were compared: one group was asked to hold the cards and deal them into one single stack as quickly as possible. Another group was asked to deal the cards into a red stack and a black stack. A third group was asked to deal the cards

into four stacks, one stack for each suit. Each participant sorted playing cards for 42 seconds. The rationale behind this paradigm was that in the first condition, participants will have to process less information than in the second condition, and similarly both conditions will require less processing than the third condition. Critically, participants were divided into two groups: the prospective group that was told that they would later be asked to estimate how long they had spent dealing cards, but that they should not count, while the retrospective group was not told anything about these temporal judgments. The results from the study by Hicks and colleagues revealed that for the prospective group there was an effect of condition: as expected, the conditions where more information had been processed seem shorter than the ones where less information had been processed. However the retrospective group did not show an effect of condition, contrary to what one would expect based on the storage size hypothesis. Hicks and colleagues argue that this might be due to the fact that participants do not have an incentive to actually memorise what they are doing and might therefore not have any reference to an amount of information stored in memory.

Therefore, Predebon (1996) investigated the effect of active versus passive processing on prospective and retrospective timing. Similar to Ornstein (1969), Predebon manipulated the number of “interval events”, which he refers to as “stimulus quantity” to assess the effect of content on duration estimation (Predebon, 1996, p. 43). In these experiments, he used sequences of visually complex patterns at different presentation rates. Participants either had to classify the stimuli into categories (active condition) or they were not required to do anything (passive condition). The results suggested that the number of interval events was positively correlated with the duration estimation for the retrospective paradigm: conditions with more “events” were rated as longer than conditions with fewer “events”. However, contrary to Hicks and colleagues’ suggestion (1976, see above), Predebon’s results did not show an effect of active versus passive processing. However, although Predebon’s task required active processing (versus passive viewing), it is questionable whether it required any aspects of memory (this will be addressed further in Chapter 4). For prospective estimates of duration

the results indicated that more events lead to shorter estimates of duration for the active condition. For the passive condition however, the results showed either a similar effect or no effect.

Extending on Hicks and colleagues' and Ornstein's findings, McClain (1983) investigated the effect of the number of stimuli presented and the level of processing required on prospective and retrospective duration judgements. In this study, McClain used words as stimuli, as she assumed that words allow for different levels of processing: graphemic conversion of a written word can be seen as shallow processing, while the semantic encoding of (aspects of) a word can be seen as deeper processing as it is more cognitively demanding. Furthermore, McClain contrasted intentional and incidental memory conditions for deep processing, meaning that participants were either instructed to memorise the stimuli, or they did not receive these instructions. Based on Ornstein's storage size model, McClain hypothesised that a greater number of words per interval should increase the storage size necessary to store the words. Based on Hicks and colleagues' findings, McClain hypothesised that the duration estimates depend on the amount of processing that is performed in an interval, which means that one would expect an effect of shallow versus deep processing. In addition, McClain predicted an interaction between the number of words and the level of processing. Her findings indicated that prospective estimates decrease when the processing demands increase. However, the level of processing did not affect retrospective estimates. The retrospective estimates were mainly affected by the number of stimuli remembered in an interval, regardless of whether the words were remembered intentionally or incidentally.

In order to further investigate what the differences are between prospective and retrospective duration judgements, Block and Zakay (1997) conducted a meta-analysis of articles on the psychology of time. In this meta-analysis, they only included articles that report research on judgements of durations over 5 seconds. Twenty research articles met all criteria and were thus included in their analysis. Their findings indicate that two variables affect retrospective judgements, but not prospective duration judgements. One of these factors is the duration length that is

under investigation. The analysis shows that the effect of subjectivity is smaller when durations are shorter. The second factor is stimulus complexity. The results show that differences in stimulus complexity lead to significant differences in retrospective duration judgements, but not in prospective duration judgements. Conversely, the results also indicate that there is one factor that only affects prospective duration judgements, namely processing difficulty. The experienced duration in a prospective paradigm decreases as the difficulty of the task increases. The authors argue that this is due to the fact that when a task is more difficult, participants have less opportunity to attend to time. These findings support the idea that attention should play a central role in models of experienced duration as measured by prospective paradigms.

In an even more recent meta-analysis, Block, Hancock and Zakay (2010) investigated the effect of ‘concurrent’ tasks during (incidental) duration encoding on duration estimation. They find that whether or not stimuli are segmented into chunks (by the experimental design rather than by the participants) greatly affects retrospective duration judgements (see below for a discussion of the definition of segmentation in this context): higher levels of segmentation predict an increase in the ratio between actual duration and subjective judgement (i.e. stimuli with more segments seem longer). Furthermore, the familiarity (in terms of continuous memory strength, cf. Yonelinas, Otten, Shaw, & Rugg, 2005; in contrast with familiarity in terms of routine, that has been shown to decrease retrospective but not prospective duration judgements (Avni-Babad & Ritov, 2003)) of the stimuli affects the duration judgements that people provide: more familiar stimuli lead to longer estimates, because the number of associations that are retrieved with familiar stimuli increases. As we will argue below, these results are compatible with cognitive models of duration estimation: higher levels of segmentation and familiarity lead to more encoded information in memory, and therefore to longer duration estimates. Additionally, the judgement method (verbal estimations lead to longer estimates than duration reproductions) and the length of the stimulus (durations longer than 60 seconds may be processed differently than shorter ones) moderate the duration judgements that people provide.

In sum, the studies above suggest that prospective duration judgements are mainly affected by attentional mechanisms and cognitive load. Retrospective duration judgements, on the other hand, are mainly affected by the encoding of the stimulus. More complex stimuli, higher levels of segmentation, higher levels of familiarity and longer durations all seem to lead to more encoded information and thus (relatively) longer duration judgements. This is supported by studies suggesting that intentional encoding affects retrospective duration judgements, as passive paradigms using incidental encoding of abstract, ('meaningless') stimuli do not seem to elicit the effects observed in studies using active memorisation of more complex stimuli.

1.2.5 The notion of events in time research

Poynter's (1989) change/segmentation model predicts that three experimental manipulations should affect duration estimates: events (the number, organisation and discreteness of events), the complexity of static stimulus patterns and the amount of cognitive load required by a concurrent task:

When the time estimate is retrospective [...] the duration judgement will depend more on the task demands of the interval. If the interval is filled with sensory stimuli, then the estimate will be based primarily on the number and discreteness of the sensory changes remembered from the interval. If the interval is empty, then the estimate will depend on the number of discrete thoughts and other organismic events the subject can remember. If the interval is filled with a processing task, then the estimate will probably depend on the number of processed items the subject can remember (Poynter, 1989, p. 314).

Evidence for the effect of memory for events comes from experiments using word lists (Block, 1974; Poynter, 1979). Both studies showed that words that are easier to remember (Poynter, 1979) or organised in a way that is easier to remember

(Block, 1974) elicit shorter duration estimates than more difficult words and word organisations. Further evidence for the role of event organisation comes from studies using auditory beeps (Adams, 1977) and flashing lights (Schiffman & Bobko, 1974; Poynter & Homa, 1983), showing that the regularity with which the beeps or lights are presented affects duration estimates: more regular patterns are estimated more accurately. Using the spacing of stimuli (target words in an array of words) as a way to manipulate the segmentation of an interval, Poynter (1983) has shown that more chunks lead to longer duration estimates. Thus, there appears to be evidence for the role of events and segmentation in duration estimation.

However, it is necessary to point out that throughout the literature, the word ‘event’ has been used to describe many different types of occurrences, not all of which are of the same nature. Particularly in time perception research, the term ‘event’ has been used to describe the presentation of any type of stimulus, including auditory tones (Vroom, 1970; Adams, 1977), flashing lights (Schiffman & Bobko, 1974; Poynter & Homa, 1983), word lists (Block, 1974; Poynter, 1979, 1983; Zakay, Tsal, Moses, & Shahar, 1994; Sahakyan & Smith, 2014), line drawings (Ornstein, 1969; Mulligan & Schiffman, 1979; Predebon, 1996) and pictures (Ornstein, 1969). Similarly, the word ‘segmentation’ has been employed to refer to the chunking or grouping of elements on many different levels. For example, Poynter (1983) investigated how duration perception changes as an effect of attending to words of a certain class (e.g. names of US presidents; “high priority events”) that are either clustered together (“unsegmented condition”) or evenly dispersed throughout a 30-word list (“segmented condition”), arguing that dispersion creates a higher number of segments, that in turn affects duration perception. However, these boundaries are induced by a top-down word monitoring process, rather than by a bottom-up perceptual process. ‘Segmented’ then means ‘aided by a task-specific segmentation or grouping mechanism’.

Note that this differs from an account that focuses on events that make up every day life: events in real life are the things that happen at various time scales, ranging from seconds to tens of minutes, that are perceived to have a beginning and an end (Zacks, 2004; Kurby & Zacks, 2008). Segmentation in this context refers

to the ability of our perceptual system to segment a stream of on-going activity into meaningful events (Kurby & Zacks, 2008). As will be argued below, Event Segmentation Theory (EST) argues that event segmentation occurs because the perceptual system is constantly monitoring the incoming information, comparing it against its predictions for upcoming information. Segmentation into discrete events then occurs at the point where the prediction error increases transiently, which leads to the subjective experience of a new event. Hence, events and event segmentation are defined in terms of subjectively identified stretches of experience, delimited by perceptually informed boundaries. This notion of events and event segmentation will be employed throughout this thesis.

Although there is some overlap between both senses of segmentation (e.g. both rely on grouping of information into discrete chunks), segmentation as defined in time perception studies is often concerned with ways to aid memory encoding, whereas segmentation in terms of EST critically predicts the perception (and, as will be argued below, subsequent memory encoding) of events. This latter form of segmentation is an automatic, bottom-up process (informed by top-down knowledge), rather than a by-product of the (experimental) cognitive task. 'Segmentation' of 'events' as employed in most time perception studies may thus be limited in describing how the human system parses experience into personal, episodic events. To the author's knowledge, the studies reported in this thesis are (among) the first to use manipulations of *natural* event segmentation to investigate the effect event perception on duration estimation from memory (see Chapter 4 for more discussion).

In sum, current research faces the challenge to pinpoint exactly what aspects of stimuli drive the differences in encoding or retrieval and therefore the expansion in remembered duration. A fruitful way of investigating the cognitive mechanisms underlying duration reconstruction thus seems to be to use a retrospective paradigm, manipulating the memory encoding or retrieval of the stimuli by changing the event organisation. This calls for an integrated approach, bringing together the literature on time perception and the literature on how we perceive and remember events in the first place in order to investigate what

aspects of events are retained and activated when we (re)construct the duration of an interval.

1.3 Memory and Events

The previous sections have given an overview of research into time perception and duration reconstruction, suggesting that retrospective duration estimations somehow rely on memory mechanisms and the amount of information that is stored. The following sections will further examine what memory mechanisms underlie remembering the content of an interval, and what characteristics of that content may account for the duration dilation effect as observed by Ornstein and others. Therefore a link is drawn between episodic memory and Event Segmentation Theory, suggesting that the way we segment and remember content may be responsible for the expansion in subjective duration.

1.3.1 Episodic memory

In the memory literature, there is a fundamental distinction between episodic and semantic memory. Formulated by Tulving (1972), episodic memory concerns the remembering of personally experienced events, whereas semantic memory concerns our memory of general facts, which were later argued to be two functionally separable memory systems (Tulving, 1983). Clinical studies of memory and amnesia have provided support for this idea. For example, based on many studies of patients with amnesia, Nielson (1958) argued that one pathway of memory is concerned with the experience of events, involving an element of time, whereas the other is centered around acquired factual knowledge, and that either can be lost without the other (Tulving, 2002). In a nutshell, episodic memory can be described in the following way:

An event happens, a person experiences it, memory traces are laid down representing the event, the past vanishes and is replaced by the present.

The memory traces of the event continue to exist in the present, they are retrieved, and the person remembers the event (Tulving, 2002, p. 19).

Tulving therefore argued that knowledge is organised in a temporal fashion in episodic memory: events can precede or succeed each other, or they can co-occur (Tulving, 1984). Hence, there is a pivotal connection between episodic memory and duration representation and reconstruction. As argued by Tulving (2002), the fact that we have a subjective sense of time, a 'self' identity and autonoetic consciousness makes it possible for us to undertake mental time travel through our subjective time. Episodic memory is concerned with "what", "where" and "when", and allows us to consciously re-experience past events, and thus, to reconstruct them and their duration.

The question that then arises is how exactly these events are remembered, and what types of information are retained in order reconstruct the event from memory. According to Shastri (2002), events are encoded in terms of their relational structure. Shastri argued that a basic representational requirement for encoding an event is to identify the specific roles that each element has in the interaction: if John gives a book to Mary on Tuesday, one needs to encode that John is the giver of the book, Mary is the recipient and that the book is the object, and that this event is anchored to a temporal location, namely Tuesday. So to encode this event, one needs to encode more than just the fact that John, Mary, the book and Tuesday are associated with each other, but one also needs to encode the relationships between these entities and concepts. This encoding provides binding at the level of the event (i.e. between all entities and concepts) but also groups together all sub-components of the event to distinguish this event from other events (Shastri, 2002).

This theory predicts that these bindings form the basis of our memory for events. Evidence from language suggests that indeed based on a sentence expressing the relationship between entities, we can reconstruct a complex event. For example, from a sentence like 'John has bought a car', we can reconstruct that an agent, John, has bought an object, a car, and that this event has already taken

place and has been completed (Shastri, 2002). Shastri argued that a sentence like the one above taps into a web of knowledge that includes semantic information about the specific entities and their attributes, knowledge about the cause and effect relations between actions and their entailed effects and an abstract, schematised representation of generic actions. An event is reconstructed by building a representation based on the relevant bindings described by the language and by activating their relevant semantic and procedural associations. Although the origin of the binding is different, these insights from language comprehension may also be applicable to event memory: an episodic memory is thought to consist of bindings that are reinstated during recall, activating their semantic associations and action schemas to reconstruct the event (Shastri, 2002).

1.3.2 Link between episodic memory and event models

This view of episodic memory is closely linked to the event perception literature. This literature is mainly concerned with how we make sense of the on-going stream of things happening by segmenting it into distinct events (Zacks & Tversky, 2001, see below). In event perception theory, relationships between the entities in an event are also pivotal (Radvansky & Zacks, 2011). Two types of relational information are of specific interest, namely *structural relations* and *linking relations*. Together, these relations cover roughly the same types of information that Shastri (2002) argued to be encoded in the episodic memory trace. Structural relations define the interrelations among the entities in the event, such as spatial relations, social relations and ownership relations. Linking relations convey for example temporal or causal relations. Temporal relations indicate when in time an event occurred, relative to other events. Causal relations are concerned with the relationship between actions and their effects, and are thought to be crucial to event model construction, as the probability of the occurrence of the cause-effect relationship may guide how event information is stored (Radvansky & Zacks, 2011). Previous research suggests that information that is contingent with the cause-effect relationships that have been established are more likely to be

interpreted as belonging to the same event, as changes in causal relationships are related with the perception of boundaries between events (Zacks, Speer, & Reynolds, 2009). Thus, the structural and linking relations of events may guide our encoding of events, as well as our perception of boundaries between events. These relations may shed light on what exactly constitutes a ‘chunk’ (cf. Ornstein) and a ‘perceived change’ (cf. Fraisse). The following section therefore reviews the relationship between the event segmentation and episodic memory in more detail.

1.3.3 Event structure

As mentioned above, the French philosopher Guyau (1890) considered time to be a purely mental construction from the events that take place, suggesting that events are crucial to our perception of time going forward. According to Zacks and Tversky, “events are the stuff of all our lives” (2001, p. 19). They argued that people segment what is going on by identifying salient boundaries (Zacks & Tversky, 2001; Zacks, 2004). This segmentation allows people to make sense of the stream of ongoing activities in terms of events. An event is defined as “a segment of time at a given location that is perceived to have a beginning and an end” (Zacks, 2004, p. 979). The following sections will review studies addressing the question what principles guide people in perceiving or deciding what constitutes such a segment of time.

One of the first studies to look into event segmentation was conducted by Newtonson (1976). In this study, participants were asked to segment movies of common activities into meaningful chunks. The boundaries of these chunks correspond to the completion of a subtask of a larger action. For example, in the activity of cooking a meal, the chopping of the onions could be seen as a subtask that needs to be completed in order for the larger task of cooking a meal to be completed. These sub-events are hierarchically structured, as a series of sub-events constitute a larger event that in turn is part of an even larger event. For example, having chopped one onion can be identified as a sub-event in the larger event of ‘chopping onions’. Newtonson found that observers are consistent and

reliable in identifying these boundaries and it has therefore been argued that these segmentations reflect an automatic, naturally occurring process that happens during perceptual processing (Zacks & Swallow, 2007).

Further evidence for the automaticity of event segmentation and hierarchical organisation of event structures comes from neuroimaging studies investigating brain activity while people passively view videos of everyday activities (Zacks, Braver, et al., 2001). After the passive viewing, participants were asked to watch the videos again and segment the on-going activities into either coarse-grained or fine-grained events. The brain activity collected during the passive viewing task was then overlaid with the event boundaries as indicated by the participant. The results showed that activity in regions in the posterior and frontal cortex increases several seconds before the indicated event boundary and peaks several seconds after the event boundary, suggesting that event segmentation is a naturally occurring, automatic process that happens during passive viewing, while participants are unaware of the segmentation task (Zacks & Swallow, 2007; Zacks, Braver, et al., 2001). Furthermore, the increase in activity is larger for the coarse boundaries than for the fine-grained event boundaries, suggesting that the brain activity reflects the hierarchical nature of event structures (Zacks & Swallow, 2007; Zacks, Braver, et al., 2001).

The fact that event segmentation seems to be an automatic process, and that observers have high agreement in identifying event boundaries suggests that there are (perceptual) characteristics that allow people to perform these segmentations. In everyday life, there are several characteristics or structures that could allow people to perceive events. Zacks (2004) argued that sensory characteristics are a likely contender to contribute to event segmentation. Especially “dynamic movement features”, such as qualitative and quantitative changes in movement seem important cues (Zacks, 2004, p. 980). He argued that people interpret these perceptual cues as relevant, even if one does not have had any experience with these cues before. He argues that for example the dimming of the lights in a theatre signals a transient change of event, even if one has never been to a theatre before.

Therefore, Zacks argued that the identification of event boundaries is a bottom-up process, as the event structure is directly extracted from sensory cues.

Secondly, Zacks claimed that knowledge structures are an underlying principle of identifying event boundaries. Zacks defined knowledge structures as “representations that capture recurring patterns of covariation” (Zacks, 2004, p. 980). Knowledge structures contain information about the intentions of an agent, which allow an observer to infer when an event starts or finishes. The fact that the event segmentation relies on these inferences means that it is a top-down process: the sensory characteristics need to be interpreted based on prior knowledge and inferences need to be made beyond the direct percept. In conclusion, Zacks argued that event segmentation relies on bottom-up processes such as the direct perception of sensory characteristics, and top-down processes such as the interpretation of physical stimuli in terms of intentions and prior knowledge.

Further evidence for the nature of event boundaries comes from studies investigating reading speed in narratives. These studies have suggested that reading time slows down at event boundaries, particularly when these are marked by temporal or causal discontinuities (Zwaan, Magliano, & Graesser, 1995). Other studies have shown that the perception of event boundaries in narratives is correlated with changes in the narrated situation, such as changes in the character’s goals, and that these behavioural observations are supported by an increase in neural activity at the event boundaries in reading (Speer, Zacks, & Reynolds, 2007). More specifically, when identifying coarse-grained segments of activity, people’s pattern of segmentation appears to be related to shifts in the focus on characters, their goals, plans and locations and the causal relations between their actions, whereas patterns of more fine-grained segmentation were more strongly related to changes in the character’s physical interaction with objects (Speer, Zacks, & Reynolds, 2004). These changes were found to predict reading times, indicating that changes in the narrated situation slow down reading times. Together, these results suggest that when readers encounter event boundaries in narratives, extra ‘processing operations’ need to be performed (Kurby & Zacks, 2008).

In order to model the interplay between perceptual input, semantic knowledge and memory in event segmentation, Zacks and colleagues have proposed Event Segmentation Theory (EST) (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). They proposed a system in which sensory input (e.g. visual or auditory) is processed to compute a series of perceptual predictions (Figure 2). These perceptual predictions are high in semantic content and contain information on for example which agents and objects are involved, what their motion trajectories are, and what their goals are. The authors claimed that these perceptual predictions are essential to organisms, as they allow for anticipation of what is going to happen rather than only reacting to what is happening, which is critical in for example avoiding predators or driving a car.

The way in which the perceptual input is processed is guided by event models, which are working memory representations of “what is happening now” (Zacks et al., 2007, p. 274). The connection between the continuous sensory input and the event models is gated such that the event models only receive information when the system is in rest. In other words, the event models are robust to the transient changes in sensory input while perceptual predictions are made, and can serve as a “constant” against which the built prediction can be checked. In turn, these event models receive information from event schemata. These are semantic memory representations that are based on prior experiences of events, and contain information on features shared between these prior events and the current event such as object movement and statistical information on sequential patterns. Thus, the event models are affected by bottom-up and top-down information: the sensory input provides information on the current perceptual environment, while the event schemata provide information based on prior knowledge.

Furthermore, the model contains an error detection mechanism that is responsible for monitoring the predictive quality of the perceptual predictions. For example, when an event becomes less predictable, the current content of the event model becomes less relevant to the perceptual predictions and thus a new event model is needed to guide perceptual processing. When there is a rapid increase in the prediction error, the representations are reset and the gate between the

sensory inputs and event models closes so that the event model can be updated, so that a new “constant” is formed. Critically, Zacks and colleagues argued that long periods of such a stable state are perceived by observers as events, and that these periods of resetting and updating are perceived as event boundaries.

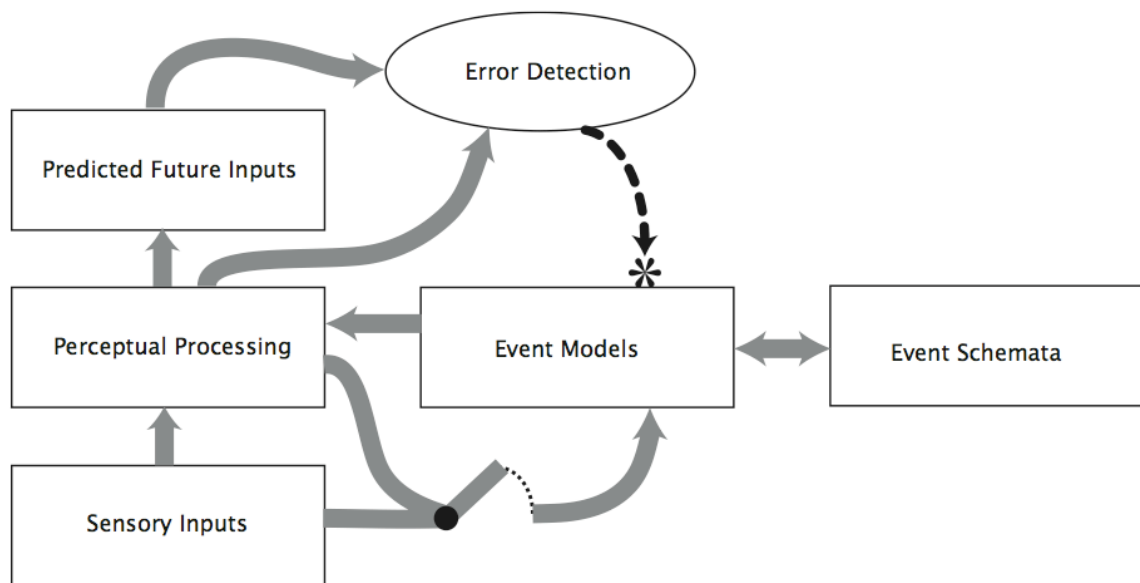


Figure 2. Diagram depicting Event Segmentation Theory. Diagram adapted from: Zacks et al., 2007, p. 2). The grey lines represent information flow between processing areas: information flows from sensory input to perceptual processing, where event models are used to guide processing in order to compute perceptual predictions. The gate in the connection between the sensory input and event models is only closed when the system is in rest. Event models are affected by event schemata that contain prior knowledge about similar events. The black lines represent the error detection mechanism, which is responsible for resetting the representations and updating the event models when there is a rapid increase in the prediction error of the perceptual predictions.

1.3.4 Event segmentation and memory formation

The segmentation of on-going events appears to form a basis for our memory of them (Zacks & Tversky, 2001). As suggested by the Event Segmentation Theory model above, working memory representations or event models are updated at event boundaries. This means that the perceptual information that is presented at that point in time is processed more extensively than perceptual information that

is not associated with an event boundary (Zacks et al., 2007). This extra processing has been argued to lead to better long-term memory representations for the information associated with the event boundary.

Evidence for this claim comes from studies investigating whether information from event boundaries is more accessible in long term memory than other perceptual information from the same stimulus. For instance, Newton & Engquist (1976) investigated recognition memory using videos of human actions. Newton and Engquist created videos of action sequences like cutting a pattern for a dress or searching for an object, and cut each video in half. Participants only studied one half of the video, watching the people in the sequence as carefully as they could. Participants then performed a recognition memory task, identifying for 72 still frames whether the frame occurred in the video that they have studied. 36 of the frames actually occurred in the studied video, half of which were selected to come from event boundaries. The other 36 frames occurred in the half of the video that the participant had not studied, so that the frames are visually similar but different in the depicted progression of events. The results showed that participants were more accurate in recognising frames that were associated with event boundaries than with a non-boundary, suggesting that information from event boundaries is more accessible.

Furthermore, the grain with which events are encoded defines the amount of information and amount of detail that can be recalled about the events. In a series of experiments using a video depicting people playing a board game, Hanson and Hirst (1989) found that finer grains of segmentation of the sequences lead to more recalled information about the events than a coarser grain of segmentation. Thus, the grain of segmentation during encoding determines the richness of subsequent recall. Similarly, disrupting the event structure by adding boundaries or disruptions in places that are not naturally event boundaries leads to impoverished memory representations. Studies investigating the effect of commercial breaks and film cuts at event boundaries show that memory for videos with breakpoints corresponding to event boundaries is as good as (if not better than) that for videos without breakpoints, whereas videos with breakpoints at non-

boundaries show impaired recall (Boltz, 1992; Schwan, Garsoffky, & Hesse, 2000; Schwan & Garsoffky, 2004, all summarised in: Zacks et al., 2007; also, see Boltz, 1995; 2005 below). Taking into account the evidence above, event boundaries appear to play a crucial role in memory encoding (see Zacks et al., 2007 for a more extensive review).

Additional evidence for the role of event segmentation in memory encoding comes from studies investigating deterioration in memory. It has been argued that event recall and recognition are impaired in older adults and adults with Alzheimer's disease, and that this is due to a deficit in encoding information (Zacks, Speer, Vettel, & Jacoby, 2006). In order to investigate the role of event segmentation in memory formation, Zacks and colleagues (2006) conducted a study in which they asked older and younger adults to segment movies showing everyday events. More specifically, they explore the effect of mild dementia of the Alzheimer's type on the relationship between event segmentation and memory by investigating the relationship between the quality of event encoding and later memory of these events. Their results indicate that the event segmentations obtained from the older adults were 'poorer' than those from the younger adults, as the segmentations were less consistent within the group. This was particularly found in the group of older adults with Alzheimer's disease. The authors attributed this effect to more general deficits in the encoding of spatiotemporal information about the events, as the older adults and patient group show impaired recognition memory for pictures of (sub-)events taken from the activities that were shown, and impaired memory for the temporal order of (sub-)events. Furthermore, their results suggested that prior knowledge about event structures (e.g. general semantic knowledge on how certain activities normally take place based on other experiences) is likely to guide event segmentation, and that therefore an attenuation of this semantic knowledge could additionally lead to poorer event segmentation. In conclusion, the findings discussed here suggest that there is a strong relationship between 'good' event segmentation and better memory for events.

1.3.5 Event segmentation and time

Although it is clear that event segmentation relies on temporal aspects of on-going events, one can also ask whether and how the segmentation of events affects our perception of time. As the perception of events and event boundaries relies on alternations between stable representations and transient resets and updates of the event models, the EST model suggests that event segmentation critically depends on change (Zacks et al., 2007, p. 277). Given the hypotheses formulated by for example Fraisse (1963) and Block and Reed (1978), who argue that important changes during an interval affect how long we estimate it to have been, one would expect a relationship between event segmentation and remembered duration.

In 1969, Ornstein already hypothesised that “in the storage of a given interval, either increasing the number of stored events or the complexity of those events will increase the size of storage, and as storage size increases the experience of duration lengthens.” (Ornstein, 1969, p. 41). He investigated the effect of ‘natural’ event segmentation on remembered duration. In his experiment, participants watched a film in which a dancer performed 26 sequenced modern dance moves for 1 minute and 40 seconds. From this film, Ornstein created three different training films: one where the dance is divided into two segments, one where the dance is divided into six elements and one in which the dance is divided into eleven elements. The content of each film was the same. As he used a retrospective paradigm, participants were naïve to the aims of the experiment, so they could not strategically remember the duration of the film. Ornstein found that the film with two segments was rated as shorter than the film with six segments, and that both were rated as shorter than the film with eleven segments. This indicates that the number of boundaries in a series of events affects duration judgements: higher numbers of boundaries lead to longer duration judgements.

Further evidence for the role of event segmentation comes from studies investigating the effect of the degree of event coherence on the memory of the duration of an event. Boltz (1995) used pieces of music and filmed narratives as stimuli, as both are hierarchically structured events of which the coherence can be

manipulated. In the filmed narrative experiment, she showed participants films with commercial breaks at natural breakpoints or non-breakpoints. Each film consisted of 20 'segments' that were marked by shifts in scene setting, and the breakpoints were either between these segments (natural breakpoints) or within a segment (non-breakpoints). Participants were asked to watch an episode of a TV series and were told that they had to perform several different tasks after watching it, unaware of the fact that they would have to provide a retrospective duration judgment of the whole episode. The results showed that the degree of predictability of the event structure (i.e. boundaries at natural breakpoints versus at non-breakpoints) affects the accuracy and reliability of the duration judgment.

In the music task, participants listened to six folk tunes that varied in rhythmic accent structure, total number of notes and, critically, temporal accents that were either compatible or incompatible with the rhythmic accent structure. Participants were asked to rate these tunes on perceptual characteristics to ensure that they would learn them, and were then surprised by a retrospective duration reproduction task. The results indicated that participants were more accurate in remembering the duration of tunes in which the temporal accents are highlighted by the rhythmic accent structure, and that this leads to less variability between subjects. This suggests that the ease of the event segmentation aids the encoding of the event in memory. Another interesting finding was that the accent-incompatible melodies were judged as being significantly longer than their accent-compatible counterparts. These findings are compatible with Ornstein's hypothesis above, as they suggest that the efficiency with which information can be encoded affects the remembered duration. This is an important link, as it illustrates that certain event structures may be easier to encode or may allow for more efficient encoding than others, and that the ease and efficiency of encoding of the event structure are predictive of the retrospectively attributed duration.

1.4 Putting together the pieces of the puzzle: duration, memory and event segmentation

So far, this thesis has argued that our memory for duration is subjective: many studies have shown that there is no such thing as absolute time perception, and that non-temporal aspects of stimuli such as their content greatly affect remembered duration. Also, this thesis has given an overview of cognitive theories of duration perception, suggesting that remembered duration critically depends on the amount of information that is remembered about a stimulus, the number of changes processed within a stimulus or the number of contextual associations that the stimulus affords. In order to remember the content of stimuli, we encode an episodic memory trace that consists of the relevant bindings between entities, semantic and procedural associations and, critically, temporal and causal relationships within and between events. These encoded bindings and relationships appear to encompass the event properties that are argued to drive event segmentation: event boundaries are perceived at points in time where changes occur in entities, their spatial locations, their goals and their interactions with objects, and in their causal and temporal relations. The fact that event boundaries are indeed predictive of how much information is encoded and can be recalled supports the idea that the way in which events are encoded in episodic memory depends on the underlying structure of the encoded events.

Linking this idea back to the cognitive models of time perception, we would expect that more encoded events boundaries lead to longer retrospective duration estimates, as processing event boundaries leads to encoding more information. Evidence from the few studies that have investigated the effect of event segmentation on retrospective duration estimation suggests that more event boundaries lead to longer duration estimates, and that the ease and efficiency with which the underlying event structure of a stimulus can be encoded modulates the duration attributed to the stimulus. However, none of these studies have manipulated the content of the stimuli in terms of their entities and their bindings, interactions and causal relations. If human duration reconstruction indeed relies

on the perceived event boundaries and the ease and efficiency with which the event structure can be encoded, we would expect that manipulating the content of stimuli in terms of the structural and linking relations within the event structure should modulate duration estimation. We would expect that when more event boundaries are identified, retrospective duration estimates are longer. Furthermore, when the underlying structure is more difficult to encode or less efficiently encoded, we would also expect an increase in estimated duration. These questions and predictions are central to this thesis, which aims to systematically investigate how event structure influences reconstructed and remembered duration.

Chapter 2

The effect of event structure on reconstructed duration

This chapter presents two experiments examining the effect of event structure on reconstructed duration. The question was how properties of the event structure such as the number of event boundaries and the efficiency with which the event can be encoded influence duration judgements based on the memory representation of what happened in the interval. Experiment 1 presented here investigated whether there is an effect of the event structure of a series of animated events on reconstructed duration, manipulating the number of event boundaries and the efficiency with which the events can be encoded by decreasing the similarity between them. As this was the first study to investigate this question using animations manipulating these aspects of event structure, a secondary aim was to identify whether indeed simple animations showing geometrical shapes interacting with each other and their environment could be used to investigate the encoding of event structure and memory for duration, and whether a learning paradigm in which participants study numerous novel animations could be used to investigate the effect of event structure on duration estimation, allowing for more generalisation across stimuli. Experiment 2 is an extension of Experiment 1, using a higher number of stimuli and a more controlled design.

2.1 Introduction

The aim of the experiments in this chapter was to investigate how the event structure of encoded stimuli influences their reconstructed duration. As discussed in Chapter 1, very few studies have looked into the relationship between event structure and duration reconstruction. Studies that have done so (e.g. Ornstein, 1969; Boltz, 1995) have restricted themselves to investigating the effect of the number and placement of imposed event boundaries, rather than natural

segmentation, on reconstructed duration. Moreover, these studies did not manipulate the content of the interval in terms of the structural and linking relations among elements in a stimulus, although intuitively, this seems like a relevant and ecologically valid manipulation: given that in real life, the same stretch of time can be filled with a monotonous action or with a variety of actions, intervals of the same clock duration can be filled with fewer or more identified event boundaries that are more or less similar to each other. Therefore, the present studies investigated the effect of the number of identified event boundaries and the similarity between them on reconstructed duration. Rather than manipulating the segmentation directly by providing artificial event boundaries, we biased people's own natural segmentation of dynamic stimuli by altering the structural and linking relations of entities within an animation.

2.1.1 Causal structure and event perception

As argued in Chapter 1, the natural identification of event boundaries critically relies on changes in the location, goals and plans of entities, their interaction with other objects and entities, and the causal relations between their actions. In particular the latter is thought to be highly informative for the identification of event boundaries, as the points in time at which causal interactions take place are often critical to the event structure: for example, a car may come to a sudden stop once it hits a tree, or a rolling ball may launch another ball when they collide (Zacks & Tversky, 2001). Thus, causal interactions are likely to inform observers about when one event ends and another begins. Furthermore, those properties of entities that are relevant to the causal structure are more likely to be stored than ones that are not directly relevant (Radvansky & Zacks, 2011). Therefore, the present study used causal interactions (e.g. colliding, bouncing) to generate event sequences with potentially different numbers of perceived event boundaries and similarity between sub-events to investigate how people's natural identification and encoding of event boundaries affects duration reconstruction.

2.1.2 Number of event boundaries

As argued in Chapter 1, the segmentation of events forms the basis for our memory of them (Zacks & Tversky, 2001). The grain of the segmentation that was used when encoding the event determines the amount of information that is recalled: more fine-grained units lead to a richer representation (Kurby & Zacks, 2008; Zacks et al., 2007). Hence, more encoded event boundaries lead to richer representations. Ornstein (1969) already has shown that the number of stimuli that have to be kept in memory affects reconstructed duration: when more stimuli need to be retrieved from memory, the interval of time seems longer. Similarly, several studies since have investigated the effect of ‘chunking’ on duration estimation by providing pre-determined stimulus chunks (e.g. Poynter, 1983; Zakay et al., 1994; see Chapter 1 for an overview). However, these studies bear little relation to perceiving and encoding a dynamic sequence of events in an unguided manner, such as on the basis of the causal structure of events. Therefore, the present studies aimed to investigate the effect of encoding and retrieving natural event segments guided by the causal contingencies perceived in the natural world rather than arbitrary chunks imposed by the experimental design. Thus, the aim of the current studies was to identify whether the same increase in duration estimate as observed by previous studies could be observed for the number of naturally identified event boundaries encoded from dynamic, causally complex events, rather than the number of predetermined chunks.

2.1.3 Similarity between sub-events

The previous chapter has outlined the idea that not only the number of event boundaries but also the ease and efficiency with which the events can be encoded affects reconstructed duration. Literature on memory encoding suggests that humans do not encode the full structure of what happened, but rather build accurate yet economical representations (Orbán, Fiser, Aslin, & Lengyel, 2008). Over subsequent exposures, similar events tend to be chunked into one representational schema, rather than stored separately (Brady, Konkle, & Alvarez,

2009). Likewise, similarity between events and in particular the repetition of similar events in variable context leads to better, more efficient memory encoding (Avrahami & Kareev, 1994; Bellezza & Young, 1989). Similarity-based mechanisms have indeed since long been recognised to play a role in learning and representing concepts or schemas. Statistical learning studies for example suggest that the repetition of a pattern or sequence in different contexts gives rise to word and speech segmentation (Gómez & Gerken, 2000). Similarly, concepts such as *dog* or *bird* are thought of as feature-based representations extracted from patterns of co-variation in our experience, which requires recognising the similarity between properties and elements in recurrent experiences (Murphy, 2002). Therefore, the efficiency with which events or series of events can be learned and encoded thus appears to be affected by the similarity between the events, sub-events and other recurrent properties. If the similarity between events or sub-events decreases, the encoding efficiency also decreases and more information needs to be stored in memory to build an accurate representation. If more encoded information indeed leads to longer duration estimates, decreasing the similarity between sub-events should increase the estimated duration. The aim of the current studies was therefore to investigate whether an event structure that can be encoded less efficiently, as there is greater dissimilarity between its sub-events, elicits longer reconstructed durations.

2.1.4 Similarity and time

Although none of the cognitive models of duration estimation have considered the effect of similarity between sub-events, the Contextual Change Model hypothesises that a complex stimuli or stimulus sequences are being judged as longer when “perhaps more varied kinds of processing were required” (Zakay & Block, 1997, p. 15). An interval may seem longer when different kinds of processing have been performed as compared to a single kind of processing because changes in the kind of processing may lead to changes in encoded contextual elements (Zakay & Block, 1997). The Contextual Change Model therefore hypothesises that retrospective

judgements rely on the retrieval of this associated contextual information (Block & Reed, 1978; Zakay & Block, 1997). Remembered duration expands as a function of the amount of contextual changes that are stored in memory and that are available to be retrieved when a duration estimate is made (Block & Reed, 1978; Zakay & Block, 1997). However, the Contextual Change framework does not determine exactly what factors affect duration estimates: changes in environmental context, mood and type of processing are proposed to influence duration judgements, yet these terms are not defined specifically, nor are their critical properties defined qualitatively or quantitatively. Thus, it remains unclear what properties of the stimulus lead to 'varied kinds of processing' and thus the expansion in reconstructed duration.

However more compelling evidence from linguistic research suggests that the similarity of associations plays a role in duration reconstruction. Coll-Florit and Gennari (2011) argued that the similarity between the sub-events of a stimulus plays a role in estimating duration. Using linguistic stimuli, they found that verb phrases (e.g. *losing money/owing money*) are judged to be longer when the phrases are associated with less similar semantic properties in long-term memory, as the semantic properties associated with the verb phrase are more distant in semantic similarity space (Coll-Florit & Gennari, 2011). For example, their findings showed that participants mainly associated *losing money* with dropping money or gambling, whereas *owing money* elicited associations with contingent concepts such as debts, debtors and loans, but also with friendships and discomfort. Coll-Florit and Gennari therefore argued that the diversity of associations for events in memory may be responsible for the duration that is attributed to them.

The effect of the dissimilarity or diversity of associations that are linked to the verbally described events is also reflected in the processing cost of these linguistic stimuli. Joergensen and Gennari (2013) used a paradigm in which they compared the interpretation of the same action verbs across different contexts that implied long and short durations. They found that the time it takes to retrieve an event that is described by a linguistic stimulus correlates with the diversity of the memory associations generated by the linguistic context. Furthermore, their

findings showed that more diverse associations are correlated with longer event durations, illustrating that the diversity of the associations varies with the implied duration. In sum, findings from linguistic research suggest that stimuli describing longer events require more cognitive effort to retrieve from memory due to the diversity and dissimilarity between their associations and contexts in which they occur. Although the conception of similarity in language and visual encoding might be different in nature (i.e. semantic versus episodic), investigating parallels across these cognitive domains could reveal general information processing mechanisms.

2.2 Experiment 1: Reconstructing duration from memory

2.2.1 Research question and aims

The aim of this study was to identify whether the event structure, and in particular the number of identified event boundaries and similarity between sub-events, affects reconstructed duration. The prediction was that more event boundaries lead to longer attributed durations, as more stored information has previously been shown to lead to longer reconstructed duration. Furthermore, the prediction was that reconstructed durations are longer when there is less similarity between the identified segments, as more information needs to be encoded for sub-events to distinguish between them in memory, and the efficiency with which a series of events can be encoded is lower, e.g., dissimilar sub-events cannot be chunked together in one schema or summary representation.

As argued above, causality is a strong predictor for where and when event boundaries are identified. This experiment therefore used the underlying causal structure of animations as a way of biasing the natural identification of event boundaries: an animation with fewer causal changes should yield fewer identified event boundaries than an animation with more causal changes. Therefore, 15 basic animations (henceforth the *basic* condition) were created that contained no or few causal changes (e.g. a ball hitting a series of squares and bouncing off them with no causal effect on the squares). The same animations were then changed to contain

repeated causal effects (i.e., changes) on the squares (e.g. the ball hitting the squares, which move away as a causal effect), inducing the identification of more event boundaries (henceforth the *numerous* condition). In order to then manipulate the efficiency with which the animation can be encoded, the animations from the *numerous* condition were altered so that different causal changes happen while the number of changes remains the same (e.g. the ball hitting the squares, which either move away or dissolve; henceforth the *dissimilar* condition). Thus, 15 triads of animations were created, each consisting of a basic, numerous and dissimilar condition (see Figure 3 for examples).

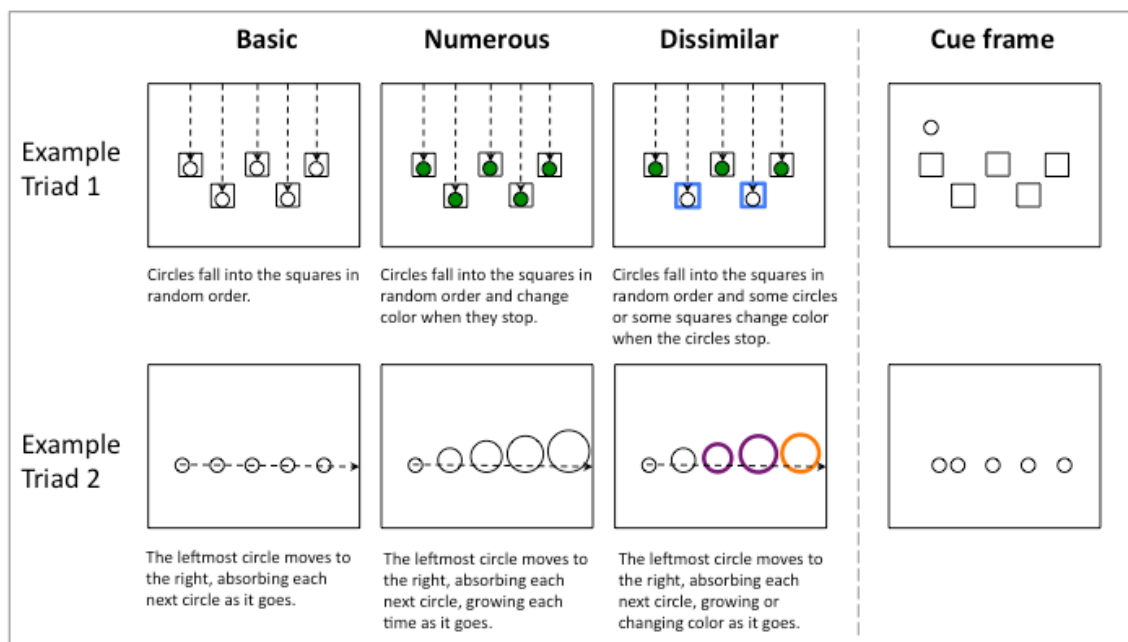


Figure 3. Example of animation triads. Dotted lines indicate the motion path of the moving shape.

2.2.2 Novel retrospective paradigm

Studies using a retrospective paradigm are far less ubiquitous than prospective studies due to the methodological challenges that the retrospective paradigm entails. An important limitation for the design of a retrospective experiment is that after asking a participant to provide a retrospective duration judgement, the participant will be aware of the aim of the experiment and will therefore strategically pay attention to duration in subsequent trials (Block & Zakay, 1997).

This means that many of the retrospective studies described above are concerned with one-off judgements of one experienced interval. For instance, in one of the experiments conducted by Ornstein (1969), each participant was presented with only one clip of modern dance. This poses a limitation for generalisation across stimuli and situations, and greatly limits the statistical power of an experimental design.

Therefore a secondary aim of this experiment was to establish an experimental paradigm that is capable of eliciting a larger number of retrospective duration judgments without the participant strategically encoding duration during stimulus exposure. To this end, this experiment used a learning paradigm in which participants first extensively studied the content of all animations over several exposures, unbeknown to them that there would later be a duration judgement task. After the study phase, participants performed a recognition memory task and a surprise duration judgement task using probed recall. If this paradigm indeed enables participants to learn the content of the animations, this could be a fruitful way of studying the effect of stimulus characteristics on duration reconstruction that can be used in future studies.

Furthermore, there has been some concern about what measure of duration is best to use in a retrospective paradigm. For instance, people might be too imprecise and variable when asked to give a verbal estimates in terms of seconds or minutes when they did not attend to temporal aspects of the stimulus, rounding their estimates to the nearest minute or half minute (Grondin, 2008). Another measure that has been used is the reproduction of an interval. However, this measure is not very suitable for longer durations and may also be highly variable across participants (Grondin, 2008). Therefore, the current experiment used a Likert scale, which allowed participants to provide a relative judgement. As participants did not intentionally attend to time while encoding the stimulus content, they are unlikely to have a representation in seconds. By using a Likert scale, the present study avoided asking participants to make an additional translation of their sense of relative duration into seconds, and might reduce variability across participants.

2.2.3 Stimulus choice

Another issue in the field of research into remembered duration is concerned with the choice of stimuli. Stimuli are often abstract and limited in their ecological validity. For instance, Ornstein (1969) used simple tones and line drawings to investigate the effect of number of chunks that have to be stored in memory on remembered duration, which are much less dynamic and semantically rich than the situations that people encounter in real life. Some studies have used more natural stimuli such as dribbling with a basketball or peddling on a bicycle (e.g. Boltz, 2005). Although these stimuli are ecologically valid, they are relatively simple and repetitive in nature; they are limited in how they reflect the vast array of dynamic changes that people encounter in the real world.

Furthermore, these highly familiar stimuli (e.g. human activities and melodies) tap into our previous experience with and knowledge of the world. Boltz (1995) for instance investigated the effect of structural coherence of melodies on remembered duration: less structurally coherent melodies are remembered as being longer than their more coherent counterparts. Similarly, Boltz (2005) used distortions in the structural coherence of visual events such as dribbling with a basketball or cycling to investigate the effect of coherence on remembered duration. However in such cases the violation of expectancy or structural coherence relies on our implicit knowledge of western musical structures and the event templates of human activities (Boltz, 1995; Schwan & Garsoffky, 2004; Zacks et al., 2007). Although these stimuli can inform us on the interplay between implicit world knowledge and duration perception, they may be less informative on how we remember and estimate duration of newly encountered events.

In sum, the current body of literature on retrospective duration judgements is lacking studies in which a large number of semantically rich stimuli is used to elicit duration estimates. Therefore, the study presented here used fifteen different triads of items. An important advance was that the stimuli used here were created to be unfamiliar and schematic enough to avoid tapping into implicit world knowledge such as schemas of human action, but to still capture dynamics of

events in the real world. For instance, the stimuli used here contained causal relationships such as pushing, hitting and colliding leading to displacement or change of properties in the numerous and dissimilar conditions. Thus, the current experiment aimed to shed light on how newly encountered events are encoded and subsequently, how their duration is reconstructed.

2.2.4 Research hypotheses

In the current experiment, participants extensively studied animations and then performed a recognition memory task and a surprise duration judgement task using probed recall. The main aim of the recognition memory task was to identify participants with poor memory for the animations, so they could be excluded from the analysis of the duration estimation data. Furthermore, previous research has suggested that detailed stimulus properties can be reactivated in recognition tasks of the kind used here (Yonelinas, 2001). In terms of event properties, we expected therefore that more event boundaries and more dissimilar sub-events would lead to longer response latencies, as more information needs to be reactivated. As outlined above, the numerous and dissimilar condition were designed to contain more event-boundaries than the basic condition, and the dissimilar condition was designed to contain more dissimilar sub-events than the basic and numerous conditions. As the stimulus pre-tests (see Table 1 below, section 2.3.3) indicated, we expected to find the largest difference in terms of response latency between the basic and dissimilar condition, as they differ most in terms of the amount of stimulus information that needs to be reactivated (i.e. number of sub-events and the similarity between them). We expected to find that the response latencies for the numerous condition would not be shorter than those for the basic condition and not longer than those for the dissimilar condition, given that the numerous condition contains more events than the basic condition, but is more internally similar than the dissimilar condition. The same logic applies to the duration judgement task: if reconstructed duration varies as a function of the event structure that needs to be reconstructed, we expected to find that the dissimilar

condition is estimated to be significantly longer than the basic condition, with the numerous condition not rated shorter than the basic condition and not longer than the dissimilar condition.

2.2.5 Stimulus pre-tests

As outlined above, the current design aimed to bias the natural segmentation of a stream of events into fewer or more sub-events that are less and more similar to each other. To ascertain that people indeed perceive a difference in the number of sub-events, ratings of the number of perceived segments were obtained. The prediction was that the numerous and dissimilar conditions are perceived to contain a higher number of segments than the basic condition. As there is no difference in number of changes between the numerous and dissimilar conditions, there was no difference predicted there. Furthermore, ratings of the relative similarity of sub-events were obtained. The prediction was that there is a difference between the basic condition on the one hand and the numerous and dissimilar conditions on the other hand in terms of the number of perceived sub-events, and a difference between the basic and numerous conditions on the one hand and dissimilar condition on the other hand in terms of perceived similarity between sub-events.

2.3 Methods

2.3.1 Participants

Fifty-two students from the University of York participated for course credit, course requirement or a small monetary reward. Ten participants were excluded because they had low accuracy scores in the recognition task (recognition accuracy $\leq .6$ in one condition) and thus contributed few data points to the condition mean given the number of stimuli. Recognition accuracy was matched across conditions (basic: $M=98\%$, $SD=7\%$; numerous: $M=94\%$, $SD=10\%$; dissimilar: $M=94\%$, $SD=9\%$). Friedman's test indicated no significant differences across conditions.

Participants had normal or corrected-to-normal vision. This experiment was approved by the Ethics Committee of the Department of Psychology of the University of York. Participants provided informed consent and were debriefed after the study.

2.3.2 Materials

Fifteen different animation triads (45 animations total) were created using Stykz 1.0.2 animation software (<http://www.stykz.net>), each item triad consisting of the three conditions outlined above. Within each item, the duration and number of frames was constant between conditions. Between the item triads, the duration and number of frames varied (average=6.9 seconds, min=3.7 seconds, max=15.6 seconds). Furthermore, different shapes and types of causal events were used (such as colour changes, shape changes, size changes, dissolving, moving and spinning) in order to prevent memory interference. For each triad, a single frame near the beginning of the animation, common to all triad members, was used as a retrieval cue.

2.3.3 Stimulus ratings

Two online questionnaire studies were conducted to collect ratings on the number of perceived event boundaries and the perceived similarity between the sub-events in the animations from an independent group of raters. A total of 60 native English-speaking participants was recruited online using Amazon Mechanical Turk. Each participant only participated in one questionnaire each (30 participants per questionnaire) and did not participate in the main experiment. Participants were given a small monetary reward for their participation.

Animations were arranged in three lists (each containing one member of each triad but all three conditions across triads; counterbalancing through a Latin square design). The same lists were used in the main experiment. The event boundary questionnaire used instructions similar to those used in event segmentation studies (Zacks, Tversky, & Iyer, 2001): participants were instructed

to indicate the number of instances in which a smallest natural and meaningful unit within the animation finished and another started. They were instructed to watch each animation several times and participants were given examples. The aim of this analysis was to investigate the stimulus characteristics, so only the by-item analysis was deemed relevant. A repeated measures ANOVA with data aggregated up to the item level as a dependent variable and condition as a repeated factor with three levels (basic, numerous and dissimilar) showed that there was a main effect of condition ($F(2,26)=8.38, p=.002, \eta_p^2 = .392$) (Table 1). One item was excluded from these analyses because the changes were too fast to allow counting. Planned pair-wise comparisons showed that the differences between the basic and numerous and basic and dissimilar conditions were highly significant (p 's < .006), confirming that the numerous and dissimilar conditions were indeed perceived as encompassing more event boundaries than the basic condition.

In the similarity questionnaire, participants were instructed to rate how similar the events within each animation were to one another on a Likert scale of 1-7 (1=not similar at all, 7=very similar). Participants were instructed to take into account the whole animation with all of its sub-events when giving a rating, and participants could watch each animation as many times as desired. Examples were provided illustrating the extreme points of the scale. As above, the aim of this analysis was to investigate the stimulus characteristics, so only the by-item analysis was deemed relevant. Results from a repeated measures ANOVA with rating data aggregated up to the item level as a dependent variable and condition as a repeated factor with three levels (basic, numerous and dissimilar) indicated that there was a main effect of condition ($F(2,28)=19.36, p<.001, \eta_p^2 = .58$) (Table 1). Planned pair-wise comparisons showed that the differences between basic and dissimilar and numerous and dissimilar were highly significant (p 's <.002). This pattern of ratings confirmed that the conditions of interest here, the basic and numerous on the one hand and dissimilar condition on the other hand, indeed differed in perceived similarity, as intended.

Table 1. Mean ratings of number of event boundaries and similarity.

Rating	Event Structure		
	Basic	Numerous	Dissimilar
Number event boundaries	4.51 (.58)	5.23 (.59)	5.31 (.53)
Sub-event similarity	5.71 (.26)	5.23 (.21)	4.23 (.18)

Standard errors in parentheses.

2.3.4 Design and procedure

Items were arranged in three lists, each containing 15 animations (5 of each condition). The items within a triad were assigned to different lists using a Latin square design. Thus participants only saw one member of each triad but several animations per condition. E-Prime2 (version 2.0.8.90) was used for stimulus presentation and response collection. Stimuli were always presented in the centre of the screen.

The experiment consisted of three tasks: a study task, a recognition task and a duration estimation task (Figure 4). In the study task, participants were instructed to study the content of the animations in association with the cue-frame for a subsequent memory task. Each trial consisted of the presentation of a cue-frame for 2 seconds, followed by the corresponding animation. After each animation, participants pressed a key on a computer keyboard to move on to the next frame-animation pair. To facilitate learning of the animation content, after all animations in a list were presented, participants were prompted to press a button when they were ready to see the animations again. Participants studied all animations in a list three times. Each study cycle presented the animations in random order.

After the study phase, participants performed a recognition memory task in which they were asked to decide whether a presented still-frame belonged to an animation that they have studied or not. As described above, cue-frames (target-frames) were obtained by extracting a frame from the beginning of each animation, so that this frame is common to all members of a triad. If recognised correctly, these target-frames should thus elicit a YES-response. In order to balance the number of YES and NO responses, foil-frames were created. As the unfolding of events is not the same for all members of an animation triad, a frame from later on in another triad member was used as a foil for the triad member that the participant had studied. For example, if the participant had studied the basic condition for a certain triad, the foil-frame would be obtained from the numerous or dissimilar condition. This means that the foil-frame presented a visually similar display, yet participants needed to employ their memory of the development of events to correctly reject the foil (i.e., a foil may look visually similar to the target, but the events have caused a slightly different outcome, or certain events have not happened in the display as depicted by the foil that have happened in the studied animation of that triad, eliciting a NO-response when correctly rejected).

Participants were instructed to indicate whether the frame belonged to one of the animations that they have studied by pressing one of two keys on a computer keyboard indicating YES or NO. In each trial, a frame appeared on the screen until participants provided a response. Inter-trial times varied randomly between 500 and 3000 ms to avoid anticipation effects. The 15 target-frames and 15 foil-frames were presented in random order during the task, thus balancing the probability of YES and NO responses across the task stimuli and avoiding response biases. From this task, response latencies to correctly accepted YES-responses were analysed. Recognition latencies longer than 2.5 standard deviations from the condition mean were considered outliers and therefore excluded from the data.

Finally, unbeknown to participants beforehand, they were asked to perform a final duration judgment task. In this task, participants judged the duration of the animations associated with the presented cue-frames on a scale of 1 to 7 using only the integers. 1 on the scale represents the shortest video seen in the study phase

(*very short*), and 7 indicates the longest video (*very long*) as compared to the other animations that they have seen. Thus, probed by the cue-frame, participants were asked to provide a duration judgment based on their memory of the animation. Participants were instructed to use the whole scale. Each trial in this task followed the presentation scheme of the recognition task, except that the trial terminated when the participant pressed a key between 1 and 7. Item presentation was randomised. Only estimates for items that were correctly recognised in the recognition memory task were analysed.

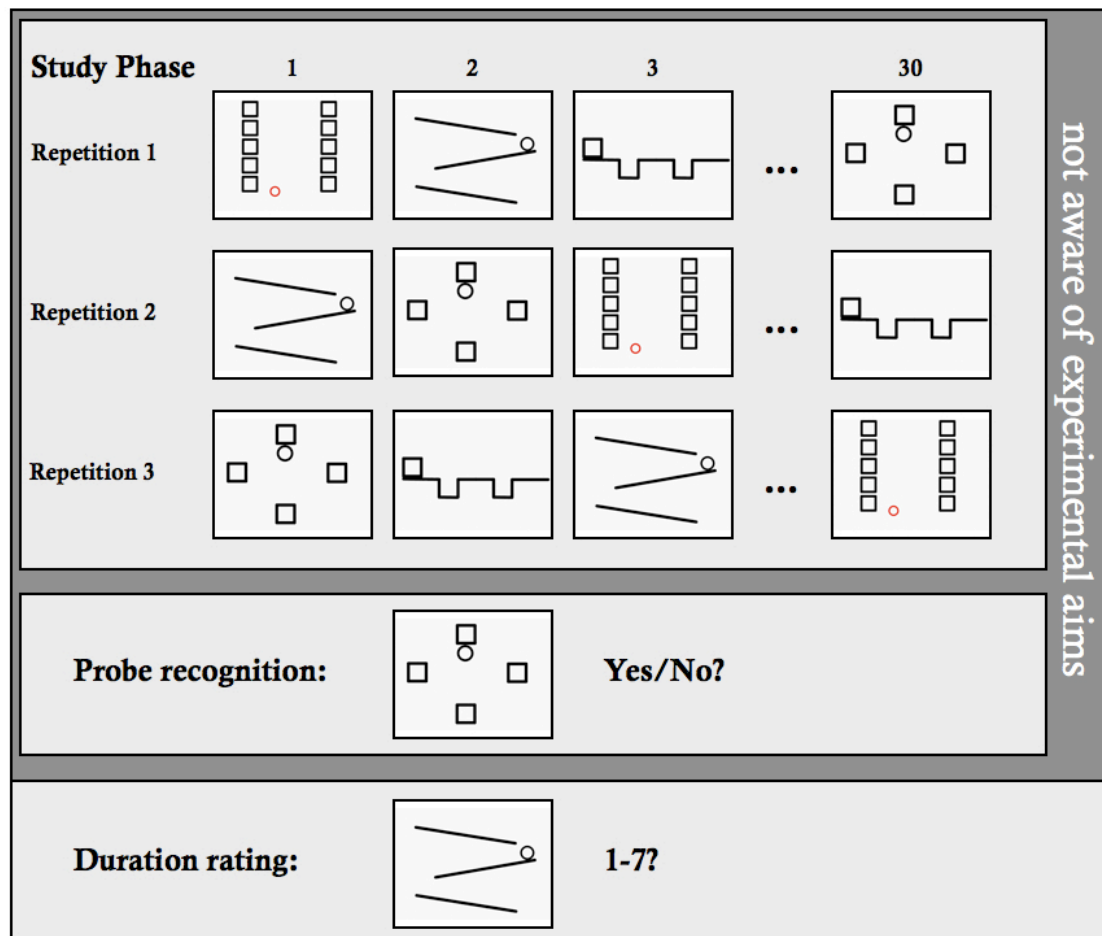


Figure 4. Schematic overview of the experiment. Participants see each animation three times, followed by a probe recognition phase. Up until this point, participants are unaware of the experimental aim (i.e. to investigate duration reconstruction). This is followed by a duration rating task.

Participants were tested in individual test booths and were told that they were participating in a memory experiment using animations to investigate how we encode memories of events. They were instructed to study short animations, and were told that they would be asked questions about their content. They were specifically instructed to pay attention to what happens in the animation, rather than to general visual characteristics such as the visual display. Furthermore, it was explained that the still frame (probe) that precedes the animation would be used later on to refer to the animation. All participants provided informed consent and were debriefed about the aims of this study after the experiment.

2.4 Results

2.4.1 Recognition memory

The prediction was that if the number of event boundaries and similarity structure determine encoding during learning and are then re-activated in cued-recognition, a main effect of condition and an increasing trend over conditions should be observed in response latencies. Repeated measures ANOVAs with recognition latencies aggregated up to the subject (F_1) or item (F_2) level as a dependent variable and condition as a repeated factor with three levels (basic, numerous, dissimilar) indicated a main effect of condition ($F_1(2, 65^*) = 13.39, p < .001, \eta_p^2 = .246$ (*Greenhouse-Geisser corrected); $F_2(2, 28) = 15.32, p < .001, \eta_p^2 = .523$), and significant linear trends ($F_1(1, 41) = 19.97, p < .001, \eta_p^2 = .32$; $F_2(1, 14) = 24.87, p < .001, \eta_p^2 = .64$). Means and standard error are given in Figure 5. These results are consistent with findings from many paired-associate and source memory studies (Yonelinas, 2001; 2002) and suggest that event properties were encoded in memory during learning and influenced the latencies of memory decisions, with more encoded information leading to increased latencies.

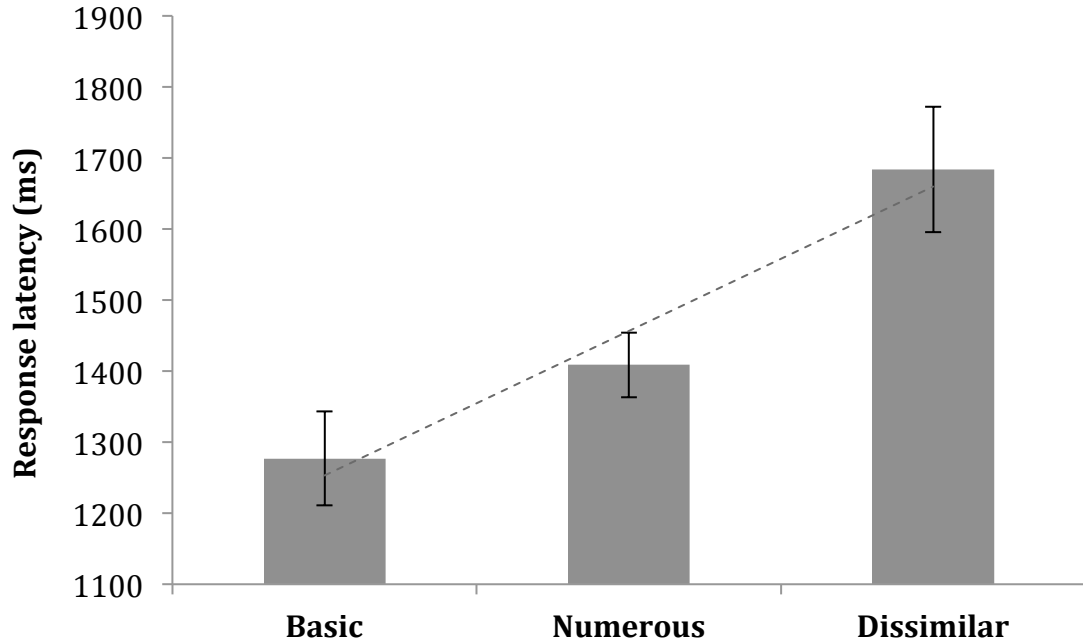


Figure 5. Response latencies over subjects. Error bars represent standard error. Trend line represents linear trend over conditions.

2.4.2 Duration ratings

The prediction was that if the number of event boundaries and similarity structure determine encoding during learning and are then employed to reconstruct duration, a main effect of condition and an increasing trend over conditions should be observed in the duration estimates, despite clock duration remaining constant. Repeated measures ANOVAs with data aggregated up to the subject (F_1) or item (F_2) level as a dependent variable and condition as a repeated factor indicated a main effect of condition ($F_1(2, 82)= 9.30, p<.001, \eta_p^2=.185$; $F_2(2, 28)=5.74, p=.008, \eta_p^2=.291$) and significant linear trends in the expected (positive) direction ($F_1(1,41)= 17.73, p<.001, \eta_p^2=.30$; $F_2(1, 14)=13.93, p=.002, \eta_p^2=.50$), indicating a significant difference between the basic and dissimilar conditions (Figure 6). Overall, these results suggest that the estimation of duration depends on the event structure people have encoded, with the dissimilar condition receiving higher estimates than the basic condition.

2.4.3 Regression results

To evaluate the independent contribution of sub-event and similarity structure on reconstructed duration, by-item hierarchical multiple regressions were used to examine the proportion of variance accounted for by sub-event and similarity scores obtained in the pre-tests over and above clock duration. Clock duration was included as a control predictor to account for the systematic variation built across triads (a factor controlled for in the analyses above). The results show that adding sub-event scores to a regression model containing clock duration as a predictor and mean ratings per item as the dependent variable significantly increased the proportion of variance accounted for (R) from .73 to .82 ($F_{\text{change}}(1, 42)=19.82$, $p<.001$). Moreover, adding similarity to this latter model significantly increased the proportion of variance accounted for from .82 to .85 ($F_{\text{change}}(1, 41)= 5.12$, $p=.03$).

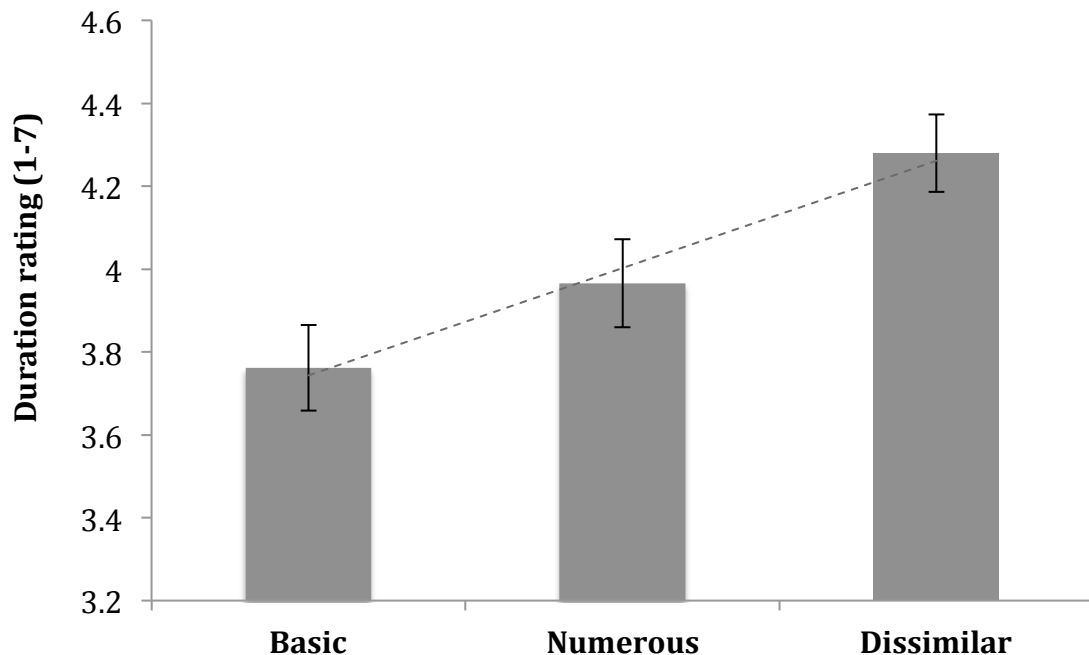


Figure 6. Duration ratings over subjects (scale 1-7). Error bars represent standard error. Trend line represents linear trend over conditions.

This pattern of significance remained regardless of the order of entry into the model (Table 2). Thus, the number of perceived sub-events and sub-event similarity played a role in duration estimation over and above clock duration, suggesting that participants estimated duration from the event structure encoded in memory. Specifically, increasing the number of sub-events in a sequence lead to increased duration estimates, whereas increasing the similarity between sub-events lead to decreased duration estimates (hence, the negative relationship in Table 2), suggesting that increased dissimilarity was related to increased duration estimations, as hypothesised.

Table 2. Regression coefficients for duration ratings. Note: * indicates $p < .05$, ** indicates $p < .001$.

	Model	<i>B</i>	<i>SE B</i>	β
1.	Constant	2.27	.27	
	Clock duration	.25	.04	.73**
2.	Constant	1.81	.25	
	Clock duration	.23	.03	.66**
	Number of sub-events	.12	.03	.39**
3.	Constant	2.83	.51	
	Clock duration	.22	.03	.64**
	Number of sub-events	.11	.03	.38**
	Similarity	-.19	.08	-.19*

2.5 Discussion

The results of the experiment presented here suggest that the encoded structure of events of the same clock duration modulates memory representations and reconstructed duration: more identified event boundaries and less similarity between sub-events lead to longer recognition latencies in recognition memory and longer duration ratings. These findings are consistent with Ornstein's

observation that the amount of encoded information affects reconstructed duration: the number of event boundaries and efficiency with which the events can be encoded modulate duration estimates. Moreover, the current findings show that indeed the similarity between the identified sub-events plays a crucial role in how the event is encoded and subsequently, how long it seems in retrospect. This will be further discussed in section 2.9.1.

2.5.1 Implications of experimental design

A secondary aim of the present study was to develop an experimental paradigm that can be used to elicit a relatively high number of retrospective duration estimates from each participant. In the current study, 15 estimates were obtained per participant. Although this number is still fairly low compared with other memory studies, it is high compared with other retrospective duration studies. Many studies used one-off judgements (e.g. Block, 1992; Hicks et al., 1976; Ornstein, 1969). To the author's knowledge, Boltz (1995; 2005) used the highest number of stimuli up until now, presenting each participant with six different stimuli (melodies or visual scenes) before asking them to provide a rating. However, one of the aims of this Thesis was to investigate the neural mechanisms that underlie duration reconstruction. In order to obtain the statistical power for an event-related functional magnetic resonance imaging (fMRI) study, more samples are necessary per participant. Therefore, another challenge was to extend the paradigm to an even higher number of duration judgements per participants. This was addressed in Experiment 2 presented below.

Furthermore, the current experiment used a Likert-scale to increase confidence in judgements and decrease variability across participants. Given the strong predictive value of clock duration on the ratings (as shown by the results from the regression analyses), rating on a Likert-scale appears to be a suitable method of obtaining duration estimates. However, even though participants were explicitly instructed to use the whole scale, it appeared to be the case that not all participants did so. This could be due to the fact that no anchoring points (i.e. the

longest and shortest animation) were provided to the participants. Therefore, in Experiment 2 presented below, anchoring points were added. Also, because the real time duration of the triads was uncontrolled, it could be the case that the spacing of actual stimulus durations affected the use of the scale: a large difference between two animations may cause clustering of animations with a smaller difference. In Experiment 2, the clock duration of the animations was therefore varied systematically.

Despite efforts to avoid stimuli resembling human action, the underlying causal structure of some of the stimuli might have tapped into top-down world knowledge rather than being indicative of the bottom-up perceptual information that is gleaned from a novel stimulus. For example, some of the events depicted by the animations were thought to be particularly Newtonian (as pointed out to us by descriptions obtained for these animations (see Chapter 3), one of the stimuli reminded some participants of a Newton's cradle) or explicitly caused by gravity (e.g. a ball rolling down steps). Therefore, in the follow up experiment, possible real-world associations were avoided by slightly altering these stimuli.

2.6 Experiment 2: Reconstructing duration from memory: extension of the paradigm

2.6.1 Research hypotheses and aims

The previous experiment has shown that recognition latencies and duration estimates displayed a main effect of condition and a linear trend over conditions. Results from the regression analyses corroborated the independent effect of the number of event boundaries and the similarity between sub-events on the duration ratings. The aim of the current experiment was to extend these findings and to improve the methodology and design. As suggested in section 2.5.1 above, the aim was to increase the number of triads in the study, thus, increasing the generalisation to other stimuli. To this end, the present study used 30 items per participant (28 triads and two anchor animations). Furthermore, the duration

judgement task here was slightly changed by providing the anchors of the scale, explicitly indicating to participants which animation was the shortest and which one the longest. This should facilitate implicit duration comparison across the stimulus set, and is analogous to providing a standard duration to which stimuli need to be compared in previous studies (Grondin, 2008). Moreover, the actual duration of triads was systematically varied so that the distance between triads is always the same (1 second) and that there are several triads per time bin (seven time bins, four animations per bin), allowing for the use of the whole scale and roughly equal amounts of ratings per point on the scale (as the use of seven time bins maps onto the seven points on the scale). The increase in the number of stimuli and the relatively small difference in actual clock duration across animations (1 sec) might elicit more variance in the data and less sensitivity to individual conditions, as less specific representations may be created during encoding of a large number of animations (30 in total). Therefore, if number of sub-events and their similarity play a role in recognition memory and duration estimation, it was predicted that at the very least, a main effect of condition and a positive linear trend over conditions should be observed. Critically, it was predicted that as the number of perceived event boundaries increases, and as the perceived similarity decreases, duration estimates should increase, as more segments and more dissimilarity between them should independently lead to more stored information.

2.7 Methods

2.7.1 Participants

Eighty-three native English-speaking students from the University of York participated for course credit or a small monetary reward. Seven participants with poor memory accuracy were excluded as they had low accuracy scores in the recognition task (recognition accuracy $\leq 50\%$ in one of the conditions or a false alarm rate above 50%). One additional participant with poor memory performance

was excluded for the purpose of counterbalancing the lists (excluding the participant with the next worst memory performance from that list; false alarm rate 39%). Correct recognition percentages did not differ significantly across conditions (basic: $M=90\%$, numerous: $M=91\%$, dissimilar: $M=87\%$; Friedman's test n.s.). Participants had normal or corrected-to-normal vision. This experiment was approved by the Ethics Committee of the Department of Psychology of the University of York. Participants provided informed consent and were debriefed after the study.

2.7.2 Materials

Twenty-eight different animation triads (84 animations total) and two anchor animations were created using Adobe Flash CS5.5, each item triad consisting of the three conditions outlined in the previous chapter. Triads varied from 3 to 9 seconds with steps of one-second increase between time bins (4 animations per bin), and the two anchor animations were 2 and 10 seconds. Within each triad, the duration and number of frames was constant between conditions. The basic animation, which depicted a stable or repeated motion of a shape, was systematically modified into the numerous condition in such a way that the speed and trajectory remained constant but an additional repeated change (in either colour, shape or displacement of another shape) was caused by its motion (see Figure 7). To obtain the dissimilar condition, the numerous condition animation was then modified so that different changes (e.g. alternating colour and shape changes) were caused by the original motion. Probes and foils were obtained as described for the previous experiment.

2.7.3 Stimulus ratings

In order to obtain stimulus ratings for similarity of the events depicted in each animation and the number of segments in each animation, two online questionnaire studies were conducted. The same list arrangements were used as in the main experiment (Latin Square). A total of 52 native English-speaking

participants were recruited using Amazon Mechanical Turk (10 - 14 participants per list). Each participant only participated in one questionnaire each and did not participate in the main experiment. Participants received a small monetary compensation for their participation.

In the segmentation questionnaire participants were asked to count the number of instances in which a smallest natural and meaningful event unit within the animation finishes and another one starts, indicating the number of event boundaries perceived in each animation. These instructions are in line with the instructions used in other segmentation studies (e.g. Newtonson & Engquist, 1976; Zacks, Tversky, et al., 2001; Zacks, 2004). Participants were given examples and were instructed to watch the animation several times. The aim of this analysis was to investigate the stimulus characteristics, so only the by-item analysis was deemed relevant. Results from a repeated measures ANOVA with rating data aggregated up to the item level as a dependent variable and condition as a repeated factor with three levels (basic, numerous and dissimilar) indicated that there was a main effect of condition ($F(2, 54) = 5.69, p = .006, \eta_p^2 = .17$) with all planned pairwise comparisons being significant ($p's < .05$) except for that between the numerous and the dissimilar conditions, as expected (see Table 3). This property reflects that the numerous and dissimilar conditions tended to have more sub-events than the basic condition.

In the similarity questionnaire participants were asked to judge how similar the events within each animation were compared to one another on a scale of 1-7 (1= not similar at all, 7 = very similar). Participants were instructed to take into account the whole animation with all of its sub-events when providing a rating. Participants were given examples that illustrated the extreme points of the scale. Participants were instructed to watch each animation as many times as they deemed necessary. As above, the aim of this analysis was to investigate the stimulus characteristics, so only the by-item analysis was deemed relevant. Results from a repeated measures ANOVA with rating data aggregated up to the item level as a dependent variable and condition as a repeated factor with three levels (basic,

numerous and dissimilar) indicated that there was a main effect of condition ($F(2, 54) = 27.05, p < .001, \eta_p^2 = .50$), and planned pair-wise comparisons were highly significant (p 's $< .001$) except for that between the basic and the numerous conditions, as expected (see Table 3). This property reflects the repetitive structure built into the basic and numerous conditions, but not the dissimilar condition.

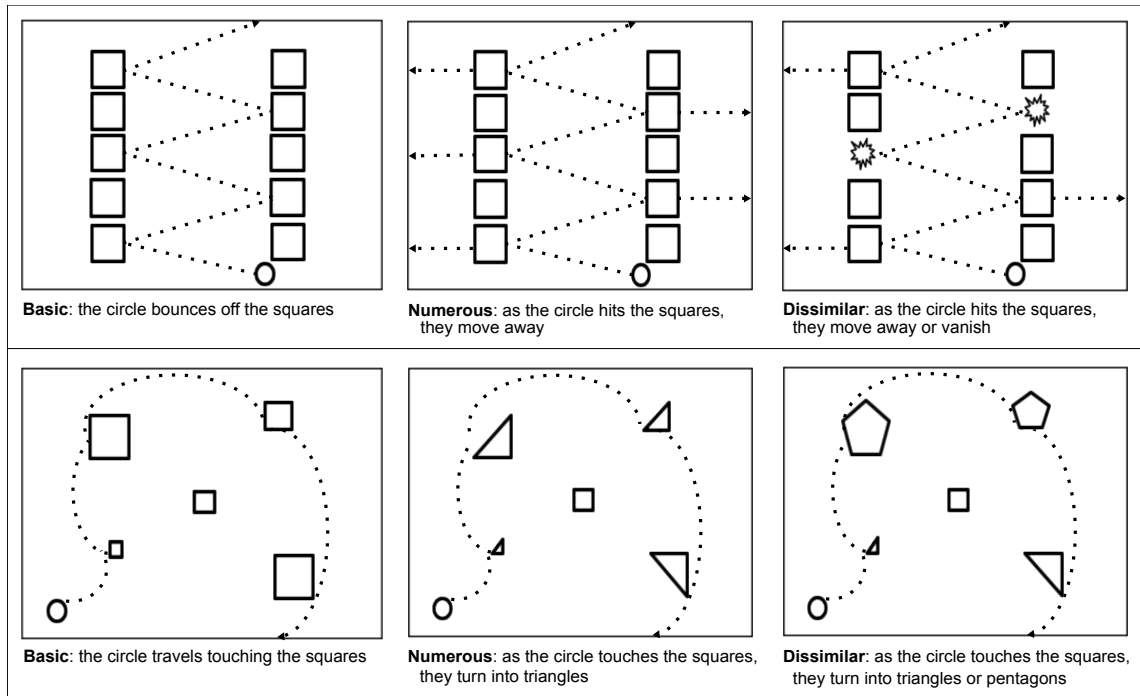


Figure 7. Example of animation triads. Dotted lines indicate the motion path of the moving shape.

Table 3. Mean ratings of number of event boundaries and similarity.

Rating	Event structure		
	Basic	Numerous	Dissimilar
Number event boundaries	4.36 (1.81)	4.74 (1.96)	4.77 (1.88)
Sub-event similarity	5.06 (1.26)	5.07 (.94)	3.89 (1.14)

Standard deviations in parentheses.

2.7.4 Design and procedure

Design and procedure were mainly the same as described for Experiment 1. In the recognition memory phase of the experiment, participants were asked to indicate whether each probe (of all 30 studied probes including the anchors and 30 visually similar foils) belonged to one of the animations that they have studied. From this task, response latencies to YES-correct responses were analysed. Responses to the anchors were not taken into account. Outliers in recognition latencies longer than 3 standard deviations from each participant's mean were excluded from these data (less than 4% of data). Analyses conducted with response latencies or their log-transforms as dependent variables yielded identical patterns of results.

A minor difference with Experiment 1 was the use of anchors. After the recognition memory task and before the duration rating task, the instructions explained to the participants that the anchor animations were the shortest and longest in the studied set and therefore outside the scale, so that the whole 1-7 scale should be used in providing the estimates. Items that were not recognised at all were excluded (incorrectly rejected and incorrectly accepted). 38 trials out of a total of 2100 trials in which a participant took longer than 10 seconds to respond were removed from the data set as these estimates were deemed unlikely to represent a confident judgment.

2.8 Results

2.8.1 Recognition memory

Based on the previous results, the prediction was that if sub-event and similarity structures determine encoding during learning and are then re-activated in cued-recognition, a linear trend across conditions should be observed in response latencies. Repeated measures ANOVAs with response latencies aggregated up to the subject (F_1) or item (F_2) level as a dependent variable revealed significant (although marginal over items) main effects ($F_1(2, 148)= 4.25, p=.01, \eta_p^2=.058$; $F_2(2, 54)=2.82, p=.068, \eta_p^2=.095$) and significant linear trends ($F_1(1, 74)=7.59, p=.007, \eta_p^2=.093, F_2(1, 27)=5.06, p=.033, \eta_p^2=.16$) (Figure 8). These results are consistent with the results from the previous chapter and many paired-associate memory findings (Yonelinas, 2001) and suggest that event properties were encoded in memory during learning and influenced memory judgments, with more encoded information leading to increased latencies.

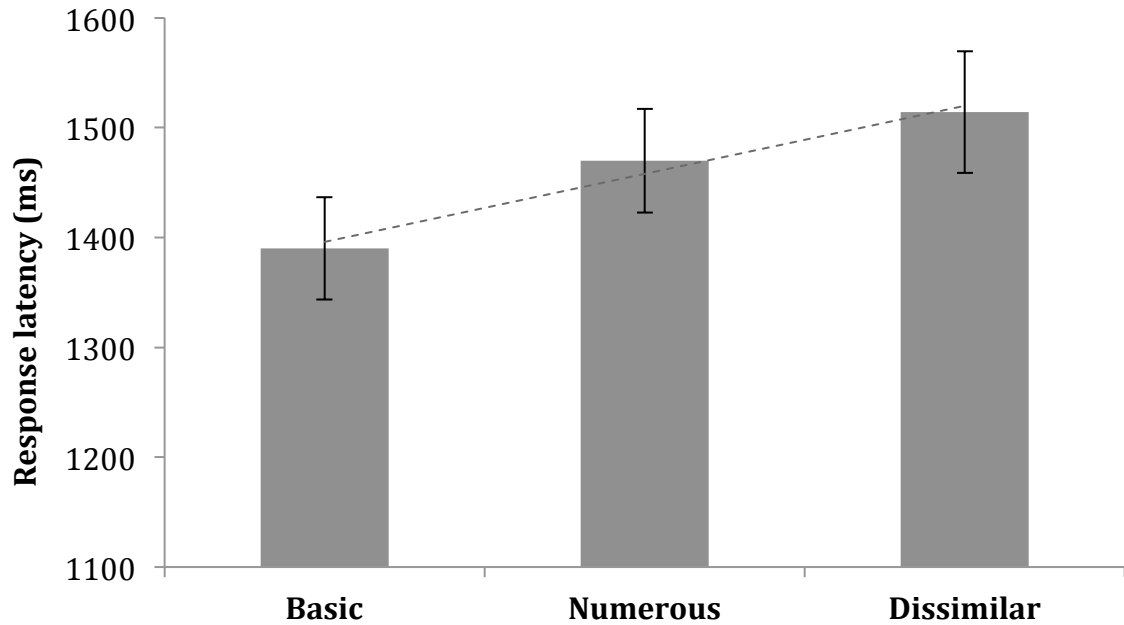


Figure 8. Response latencies over subjects. Error bars represent standard error. Trend line represents linear trend over conditions.

2.8.2 Duration ratings

Based on the previous results, the prediction was that estimated duration would increase across conditions, despite clock duration remaining constant. Repeated measures ANOVAs with duration rating data aggregated up to the subject (F_1) or item (F_2) level as a dependent variable revealed main effects of condition ($F_1(2, 148)=5.41, p=.005, \eta_p^2=.07$; $F_2(2, 54)=3.16, p=.05, \eta_p^2=.11$) and significant linear trends ($F_1(1, 74)= 11.88, p=.001, \eta_p^2=.14$; $F_2(1, 27)= 5.86, p= .02, \eta_p^2 = .18$) (Figure 9). These results replicate the findings from the previous study and suggest that event properties encoded in memory are retrieved during duration estimation, with increasing duration attributed to events of same clock duration as a function of condition.

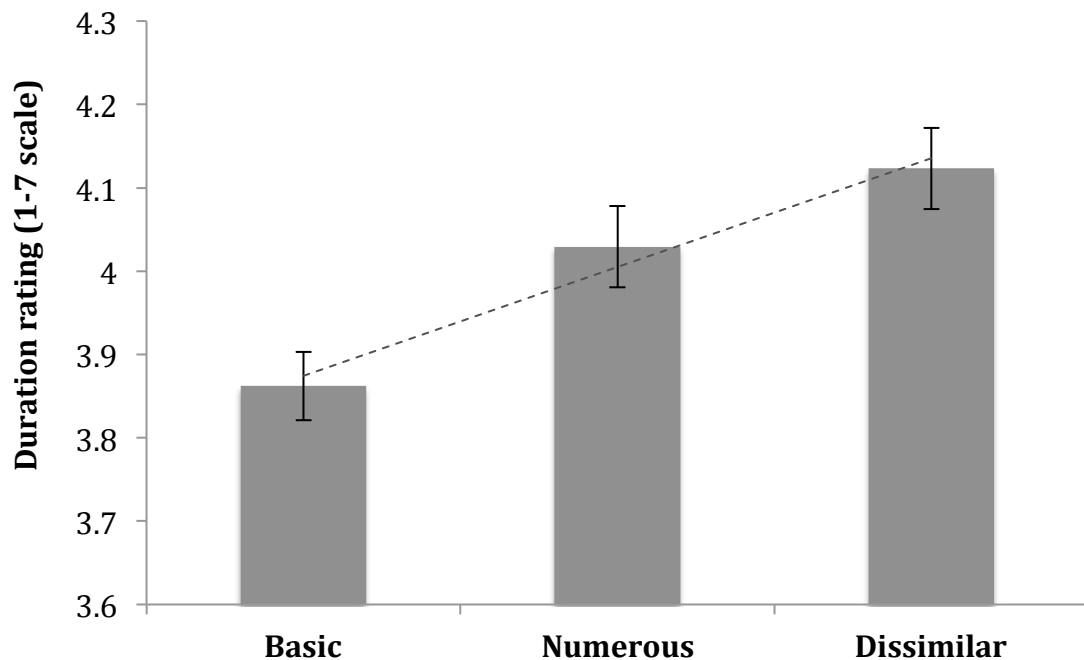


Figure 9. Duration ratings over subjects (scale 1-7). Error bars represent standard error. Trend line represents linear trend over conditions.

2.8.3 Regression results

To further evaluate the independent contribution of sub-event and similarity structure on the duration ratings, by-item hierarchical multiple regressions were conducted to examine the proportion of variance accounted for by the number of identified event boundaries and perceived similarity as obtained in the pre-tests over and above clock duration. Clock duration was included as a control predictor to account for the systematic variation built across triads (a factor controlled for in the analyses above). The results showed that adding sub-event scores to a regression model containing clock duration as a predictor and mean ratings per item as the dependent variable significantly increased the proportion of variance accounted for (R) from .69 to .74 ($F_{\text{change}}(1, 82) = 12.77, p = .001$). Moreover, adding similarity to this latter model significantly increased the proportion of variance accounted for from .74 to .77 ($F_{\text{change}}(1, 80) = 8.53, p = .005$) (see Table 4). This pattern of significance remained regardless of the order of entry into the model. Thus, the number of perceived sub-events and sub-event similarity play a role in duration estimations over and above clock duration, suggesting that participants estimate duration from the event structure encoded in memory. Specifically, increasing the number of sub-events in a sequence leads to increased duration estimates, whereas increasing the similarity between sub-events leads to decreased duration estimates (hence, the negative relationship in Table 4), suggesting that as hypothesised, increased dissimilarity is related to increased duration estimations. These findings replicate the findings from the previous experiment.

Table 4. Regression coefficients for duration ratings. Note: * indicates $p \leq .005$

	Model	<i>B</i>	<i>SE B</i>	β
1.	Constant	2.14	.23	
	Clock duration	.31	.04	.69*
2.	Constant	1.98	.22	
	Clock duration	.20	.05	.45*
	Number of sub-events	.18	.05	.36*
3.	Constant	2.77	.34	
	Clock duration	.19	.04	.42*
	Number of sub-events	.18	.05	.36*
	Similarity	-.15	.05	-.21*

2.9 Discussion

The aim of the present experiment was to extend, replicate and generalise the findings from Experiment 1, using an improved and extended design. The findings here indeed suggest that the encoded sub-event and similarity structure for events of the same clock duration modulated memory representations and duration estimation, as more sub-events and less similarity between them lead to longer response latencies and duration estimates. Therefore, the paradigm could be used successfully to elicit higher numbers of duration estimates.

2.9.1 General discussion

Overall, the results of these experiments suggest that memory encoding mechanisms modulate our reconstruction of events and event durations. The recognition memory results suggest that the event structure that has been learned for each animation is reactivated to an extent consistent with the segmentation and similarity manipulations over conditions. When the encoded representation

contains more segments and is less succinct due to greater dissimilarity, it takes longer to identify a given frame as belonging to the studied animation.

Interestingly, at the item level, segmentation and similarity ratings only predicted estimated durations and not recognition memory latencies (the latter was explored but not reported here due to the lack of significant results). This is consistent with a difference in tasks demands: in the memory task, participants should respond as quickly as possible, and they do not need to reconstruct the animations in their entirety to make a judgement (see 7.2.4 for a more extensive discussion). In the duration estimation tasks, however, they must retrieve as much of the animation as possible to be able to evaluate its duration. Therefore, a deeper reconstruction of the event takes place such that segmentation and similarity play a critical role in the duration estimation task. However, the increase in response latencies for the recognition memory task suggests that to some extent, recollection does contribute to item recognition as well, as familiarity-based recognition alone does not explain the pattern of results (note that as per design, there was no difference in recognition accuracy between conditions). These findings are in line with previous meta-analyses of recognition memory studies, suggesting that although recognition memory tasks may not critically depend on the reconstruction of events, recollection of some event detail may still contribute to the recognition process (Yonelinas, 2002). An increase in response latencies thus suggests that more event detail may be recollected about conditions with more sub-events and less similarity between them.

The findings from Experiment 1 and 2 show that indeed the similarity between the identified sub-events plays a crucial role in how an event is remembered and subsequently, how long it seems in retrospect. These findings suggest that a model in which reconstructed duration is regarded as an index of the number of perceived changes (e.g. Fraisse, 1963) is overly simplistic. Over and above the number of changes, the results here show an effect of the similarity between these changes. This finding however, can be made consistent with a model of contextual change (Block & Reed, 1978), provided that the similarity between sub-events is construed as some sort of contextual change. It is possible for

example, that for similar repeating sub-events, the ‘context’ of the sub-events (preceding and subsequent events) does not change, whereas they do in the dissimilar condition. The outcomes of the present experiment thus highlight two properties of events that contribute to the ‘complexity’ of the stimulus and therefore to the amount of ‘varied processing’ in terms of the contextual change model, namely the number of event boundaries and the similarity structure of the event.

The role of similarity in duration attribution was already highlighted by linguistic research indicating that described events that have more sub-events and a greater variety of associated events in memory are judged as longer (Joergensen, 2008; Coll-Florit & Gennari, 2011). However, in language research, it is very difficult to tease apart the effect of the number of sub-events from the actual typical duration of an event. For example, *building a cabinet* is associated with more sub-events and rated longer than *opening a cabinet*, which is not entirely surprising given that *opening* is an instantaneous event whereas *building* is a durative event. The use of visual events in the present study is an important advance, as it provides the opportunity to keep the actual duration identical between stimuli while systematically varying the number of perceived sub-events and the similarity between them. A follow-up experiment (Experiment 4) presented in the next chapter investigated whether the same pattern of results can be observed for descriptions of events that are equally long in clock duration but differ in the number of sub-events and similarity between them. In sum, the findings from the current study corroborate the suggestions from language research by showing that when event duration is kept constant, an effect of associated sub-events and similarity between them on duration judgements is observed.

This is not to say that the conception of similarity in language and visual encoding must therefore be of the same nature. In semantic memory, (definitions of) events are stored in association with the multiple situations and contexts in which these events can occur, creating a measure of contextual diversity or dissimilarity as measured by Coll-Florit and Gennari (2011). The episodic representations encoded from the visual stimuli presented here are based on the

event properties gleaned from a few exposures of the same stimuli. However, within an animation, each event boundary occurs in the context of other event boundaries. In the dissimilar condition, there is thus greater contextual diversity than in the other conditions. This again points out the contribution of the current findings to models of cognitive timing that are based on detecting change, namely that contextual diversity may arise from the dissimilarity between sub-events as well as changes in processing or environmental context. Nevertheless, there may be limitations with regard to the contextual diversity manipulated by the current experiment. These limitations and avenues for overcoming these limitations are discussed in Chapter 7.

Furthermore, the present findings are consistent with literature on memory encoding and event segmentation in general. According to Orbán and colleagues (2008), humans extract chunks from the on-going stream of information in order to build an accurate yet economical representation of what happened, rather than encoding the full structure or sensory experience. The findings here are consistent with the claim that more information is encoded when there are more identified segments and less similarity is perceived between them. These findings confirm that manipulations in event structure can drive natural event segmentation and may thereby modulate the amount of stored information.

2.10 Conclusion

In sum, the results presented here suggest that the number of identified event boundaries and the similarity between sub-events are important aspects of our memory representations that modulate our reconstruction of the time passed: more identified event boundaries and less similar sub-events lead to longer response latencies in recognition memory and longer duration estimates. Since more event boundaries and less similarity between event components lead to more stored information (as suggested by the pattern observed in the recognition memory task), the more information that has been stored in association with an event, the longer the event is estimated to be in retrospect.

Chapter 3

The effect of event structure on reconstructed duration based on linguistic descriptions of events.

3.1 Introduction

This chapter presents two experiments in which the relationship between descriptions of visual events and duration estimates was explored. More specifically, the experiments presented here focused on similarities and differences between visual and verbal encoding, and the relationship between verbal encoding and duration estimates. The first study here presents an exploration of how visually presented stimuli are described verbally from memory. These descriptions were analysed qualitatively, investigating whether linguistic characteristics correlate with event qualities and duration estimates. The second experiment presented here aimed to investigate whether matched descriptions of equally long events with different underlying event structures lead to different duration estimates. Together, these studies aimed to identify whether properties of verbal descriptions correlate with duration estimates, providing a comparison between how people estimate duration based on language and vision, suggesting similarities and differences between the two.

3.2 Memory for duration and language

When for instance an eyewitness gives a witness testimony, he or she is likely to provide a verbal description of the unfolding of events that happened. This description of what the witness did or saw could be critical to building a mental representation of the past events in the mind of others, including the judge or jury. Based on the mental representations that people (re)create from language, all sorts of decisions and inferences may be made that may for instance be critical for the outcome of a trial. Some of these inferences may include duration representations, and thus it is important to understand whether verbal reconstructions affect the

way people represent events and their durations. Hence, the relationship between how events are described and how duration is attributed to these events is a critical one, as differences in the estimation of event duration can be challenging for the evaluation of eyewitness testimonies (Burt, 1999). Although it is not a broadly studied area, previous studies have hinted at a relationship between the way events are described and the duration attributed to them. The following sections will provide an overview of previous research into the relationship between verbal description and duration estimation.

3.2.1 Retrospective interference through linguistic post-event information

There is evidence that the wording that is used to refer to an event or series of events affects people's estimate of duration. This is known as post-event information, as it is concerned with how the event is described after it occurred, causing retrospective interference (Loftus & Pickrell, 1995). Interestingly, this information percolates into the encoded representation, altering the reconstructed duration. One of the main avenues of research into this area is concerned with the speed that is implied by the language that is used to probe a witness or participant to give a time- or speed-related estimate. For instance, Loftus and Palmer (1974) conducted a study in which participants were shown films of traffic accidents and were asked how fast the cars were going before they smashed into each other. The wording of these lead questions was manipulated: 'smashed into each other' was replaced with a different phrase (collided, bumped, hit, contacted) to investigate whether different verbs imply different speeds. Their results showed that indeed different verbs elicit different speed estimates.

Furthermore, studies have shown that the speed implied by post-event wording affects duration estimation. In a study by Burt and Popple (1996), participants witnessed a staged event of a confederate bursting into a lecture theatre. Two weeks after the incident, participants were asked about the duration of the event. The phrasing of the lead question was again manipulated: participants

were asked how long (in seconds) it took the confederate to 'run', 'walk', or 'pass' through the lecture theatre. Participants in the 'run' condition estimated the duration as being shorter than the participants in the 'walk' condition, suggesting that duration estimates are affected by post-event information such as the phrasing of the lead question.

In sum, post-event information can influence duration estimation. When for example a higher speed is implied by a lead question, the duration seems shorter. This means that the language used to refer back to these events or summarise the events percolates into the actual memory of the unfolding of the events. So even when a person witnessed the events themselves, the wording that was used to refer to the events can affect their remembered duration, suggesting that verbal descriptions of the same event can lead to different duration estimates.

3.2.2 Retrospective interference through verbal description

Besides the research on the phrasing of lead questions, research has also looked into whether describing events in a certain way retrospectively interferes with remembered duration. Burt (1999) investigated whether describing a series of events with fewer or more words implying more or less action has an effect on remembered duration. He instructed participants to watch a video of a robbery and to provide a description of the robbery from memory. After participants had completed their description, they were asked to retrospectively provide a duration estimate of the original events in the video. An analysis of the content in terms of the total number of words and total number of action words showed that although there is no correlation between the total number of words and the estimated duration, the total number of action words was negatively correlated with the duration estimates, suggesting that a higher number of action words leads to shorter duration estimates. To some extent, this finding appears to be counter-intuitive: based on the findings from Chapter 2, one might have expected that a higher number of action words would be associated with more events and thus, longer durations. However, an issue with these findings is that Burt did not obtain

speed ratings for the verbs used in the descriptions. Thus, it could be the case that, as expected by the effect of different implied speeds in lead questions, qualitatively different descriptions of speed may lead to different duration estimates. Furthermore, Burt did not take into consideration the number of different action words used: following the findings presented in Chapter 2, one could hypothesise that a decrease in similarity between the action words used leads to longer duration estimates.

The finding that there is no relationship between the total number of words and the duration rating is surprising, as one would have expected a positive correlation based on Ornstein's hypothesis that more information leads to longer duration estimates from memory and that this should be reflected in the number of words. Pedersen and Wright (2002) argued that there may be no – or only a very minor – correlation between the way the events are verbalised and duration judgements. They argued that there have been case studies showing that in certain types of amnesia there is a dissociation between memory and temporal judgements (e.g. Sirigu & Grafman, 1996) and therefore question the direct relationship between language and duration estimation. They criticised Burt's research because of the correlational nature of his findings: one cannot distinguish whether the duration ratings are modulated through retrospective interference by the way in which the events are described, or whether the differences arise from differences in memory representation.

In a series of experiments that aimed to scrutinise the effect of retrospective interference, Pedersen and Wright (2002) therefore revisited the relationship between the way events are described and duration estimates by focusing on whether descriptions written in different styles (tabloid, personal or police) lead to different duration estimates. Participants witnessed an incident during a lecture (similar to Burt & Popple, 1996) and were afterwards asked to write a description of the events in a certain style and then provide a duration judgement. Thus, in line with studies looking at the effect of retrospective interference, this study investigated the effect of writing a description in a certain genre on the writer's own remembered duration. Based on Ornstein's Storage Size hypothesis, Pedersen

and Wright expected to find that when a longer description has been written, the duration of the original events should be estimated to be longer. Furthermore, based on Burt's (1999) findings, they expected that describing the events using verbs that imply a higher degree of action should lead to shorter duration estimates. Contrary to Burt, their findings showed no significant differences between duration judgements following different writing styles when comparing between conditions (between participants), even though different styles lead to different numbers of words, verbs and implied action. These findings suggested that describing events in a certain way may not alter the participant's own remembered duration.

However, results from a second study conducted by Pedersen and Wright (2002) using videos of crimes from television programs suggested that in a counterbalanced within-participant design, there might be a relationship between writing style and duration estimates. Moreover, although Pedersen and Wright claimed that these correlations are small and should therefore be interpreted with caution, for personal accounts, duration estimates appeared to be correlated with the number of words ($r=.41$). As these are personal (i.e., episodic) accounts (not altered by a genre effect) this finding may reflect a relationship between more encoded information and longer duration estimates (rather than retrospective interference). Furthermore, for police statement-style descriptions, a negative correlation was found between the amount of implied action and duration estimates ($r=-.39$; higher implied action leads to shorter duration estimates, e.g. 'walk in' vs. 'rush in'). Although these correlations did not reach statistical significance after corrections for multiple comparisons, they are in line with findings from previous studies (e.g. Burt & Popple, 1996; Burt, 1999) suggesting that describing events with a certain implied speed may alter the writer's remembered duration.

3.2.3 Duration estimation based on verbal description

In the studies described above, participants experienced the actual duration of the events that they would later describe. That is, they actually witnessed the unfolding of the events. This is not the case when for instance a juror has to reconstruct an event representation and its associated duration from a description provided by a witness. To investigate the effect of wording on duration judgements, Burt (1999) conducted an experiment in which he systematically varied the implied speed of a series of events in a robbery as described by a witness, and asked participants to read the account and estimate the time the witness would have had to view the robber. In this case, participants had to base their duration estimates solely on the provided description, without witnessing the actual events. Findings from this study suggested that events that are described using 'fast' words are rated as being shorter than the same events being described using 'slow' words.

Using the personal, tabloid-style and police statement-style descriptions as obtained from the experiment explained above, Pedersen and Wright (2002) asked a new group of participants to read these descriptions and provided a duration rating based on the descriptions only. Their results showed a small but significant effect of writing style on the duration estimates, although they did not see a clear relationship between this effect and any of their linguistic measures (number of words, number of verbs and the action intensity of each verb (e.g. walking versus rushing)). In a discussion of these findings, Burt (2002) argued that these small but significant effects should not be diminished, as even small modulatory effects of the content of descriptions on duration estimates could have major implications in real life, for example in court. Furthermore, as the present chapter will show, verbal descriptions could give rise to representations of the described event structure, going beyond the individual words used in the description. Therefore, Experiment 3 and 4 are concerned with investigating the relationship between descriptions in terms of their characteristics in terms of words and verbs, and the mental representations of events employed in duration estimates.

3.2.4 Research aims and questions

As outlined above, there is evidence for a relationship between the way events are described and their estimated duration: different descriptions of the same events elicit different duration estimates. However, there is little agreement on what aspects of descriptions give rise to these differences. Therefore, one of the aims of the experiments presented here was to further investigate the relationship between characteristics of verbal descriptions and duration estimates, focusing on the relationship between the number of words and verbs used to describe a stimulus and the event properties of that stimulus. Experiment 3 therefore explored whether events of the same clock duration but with a different underlying event structure elicit descriptions from memory that are different in terms of the number of words, number of verbs and number of different verbs, and whether these properties are correlated with duration ratings and measures of underlying event structure.

The second aim of the experiments presented here was to identify whether the event properties of the described stimulus affect duration estimates in the same way that they affect duration estimates from memory as shown in Experiment 1 and 2. Therefore, Experiment 4 investigated whether length- and speed-matched descriptions of events with different event structures but with the same clock duration give rise to different duration estimates, and whether these estimates could be explained by properties of the description and properties of the underlying events. Together, these studies aimed to address whether mental representations of events that arise from language incorporate the same event characteristics as those encoded from visual events, shedding light on how event representations arise from language and how we employ this representation to inform us about the duration of events.

Because the aim of these studies was to investigate the mechanisms that underlie event representation and memory encoding, we aimed to control for the influence of previous schema knowledge. For instance, we have top-down predictions about the unfolding of typical events based on event schemas that we

have built over previous experience that could inform us when we estimate duration (Avni-Babad & Ritov, 2003; Zacks et al., 2007). To control for influences of previous schema knowledge, unfamiliar stimuli were used here that are abstract enough to avoid tapping into top-down knowledge, but do reflect the same kind of dynamics that one would find in the real world.

3.3 Experiment 3: Exploration of the effect of event structure on naturalistic descriptions of events.

3.3.1 Research hypotheses and aims

The first study presented here explored how people describe events of the same duration but with a different underlying event structure from memory. In particular, the study examined whether linguistic characteristics correlate with duration ratings. Chapter 2 argued that more information needs to be stored and retrieved for events that contain more identified sub-events with less similarity between them. Therefore, the present study aimed to identify whether characteristics of descriptions reflect these differences in event structure. Three linguistic measures were investigated. Firstly, the total number of words used to describe the events was investigated, as the number of words could be seen as a crude measure of the amount of information (cf. Ornstein, 1969; Pedersen & Wright, 2002). However, given that not all words bear equal amounts of information about events, in addition, the total number of verbs was also investigated, as these might be expected to reflect actions and events in the animations (cf. Pedersen & Wright, 2002). Finally, the total number of different verbs was investigated, as the use of different verbs might indicate more dissimilar events, suggesting a richer representation in memory (see Chapter 2). Patterns of these three measures were explored across conditions, and their relationship with the number of identified segments, the perceived similarity between sub-events and duration ratings was examined. To ensure that participants had a comparable level of encoding as those in Experiment 1 before providing descriptions, this study

used the same learning paradigm and stimuli as Experiment 1. Thus, prior to providing a description of the animations, participants studied 15 animations of geometric shapes moving and interacting with each other, and performed a recognition memory task and a surprise duration rating task.

3.4 Methods

3.4.1 Participants

Thirty-seven native English-speaking students from the University of York participated in this study. The experiment was approved by the Ethics Committee of the Department of Psychology of the University of York. Participants provided informed consent and were debriefed after the study.

3.4.2 Materials

The same materials were used as in Experiment 1 presented in Chapter 2. The description of the materials can be found in section 2.3.2.

3.4.3 Design and procedure

This experiment had the structure of Experiment 1 described in Chapter 2. This means that participants studied 15 animations three times, performed a memory recognition task and provided duration ratings of the animations that they studied. Participants were then presented with a webpage that displayed all of the probes of the animations that they had studied, each accompanied by a textbox. Participants were instructed to provide a description of each animation based on their memory of its content. They were told that they could use as many words and as much detail as they felt was natural and necessary.

Data were then coded for content. For each of the descriptions, the number of words was counted. Furthermore, the total number of verbs was counted, excluding verbs that were only used to mark tense (e.g. 'is' or 'has' in a past perfect context). Thus this measure included counts of tokens of the same verb.

Furthermore, the number of different verbs was counted, counting each occurrence of the same verb stem as the same verb (e.g. 'running' and 'ran' were counted as the same verb, same for e.g. 'is' and 'be'). Meta-comments (e.g. 'I think') were excluded from the analyses. The same ratings for event properties of the stimuli were used as in Experiment 1. For correlation analyses between linguistic properties and duration estimations, we used the estimation data from Experiment 1, as the ratings of this experiment were not recorded correctly due to technical issues.

3.5 Results

The aim of this experiment was to explore how people describe events from memory and whether properties of these descriptions correlate with event properties and duration estimates. The aim of the analyses presented here was to investigate properties of the item descriptions (rather than generalisation to the subject population level), so only the by-item analyses were deemed relevant. The focus was on three measures: the number of words, the number of verbs and the number of different verbs, as differences in these measures over conditions may be indicative of differences in the amount of stored information. The assumption that these measures are related with event properties was tested, and their relationship with duration estimates was interrogated.

3.5.1 Number of words

The results suggested that the number of words used in the description increased over conditions (Figure 10). Results from a repeated measures ANOVA with the number of words aggregated up to the item level as a dependent variable and condition as a repeated factor with three levels (basic, numerous and dissimilar) indicated that there was a main effect of condition ($F(1.39,19.45)=14.29$, $p < .001$, $\eta_p^2 = .51$). Furthermore, there was a significant linear trend over conditions ($F(1,14)=20.15$, $p=.001$, $\eta_p^2 = .59$).

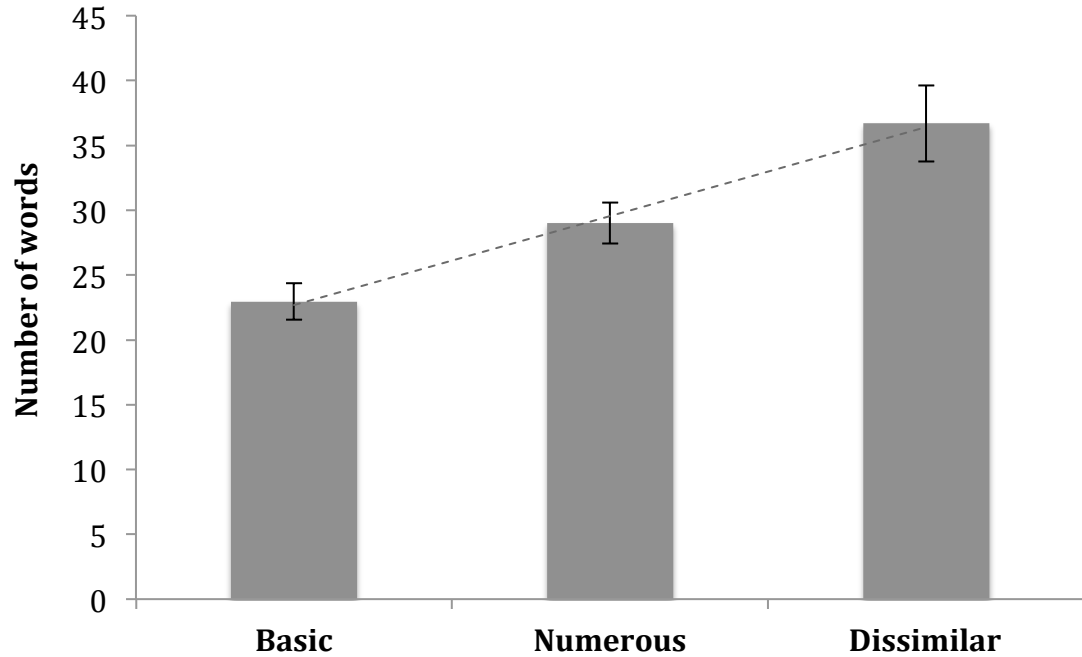


Figure 10. Average number of words per item per condition. Error bars represent standard error. Trend line represents linear trend over conditions.

3.5.2 Number of verbs

The results suggested that the number of verbs used in the description also increased over conditions (Figure 11). Results from a repeated measures ANOVA with the number of verbs aggregated up to the item level as a dependent variable and condition as a repeated factor with three levels (basic, numerous and dissimilar) indicated that there was a main effect of condition ($F(1,14)=22.76$, $p<.001$, $\eta_p^2=.62$ (*Greenhouse-Geisser corrected)). Furthermore, there was a significant linear trend over conditions ($F(1,14)=34.09$, $p<.001$, $\eta_p^2=.71$).

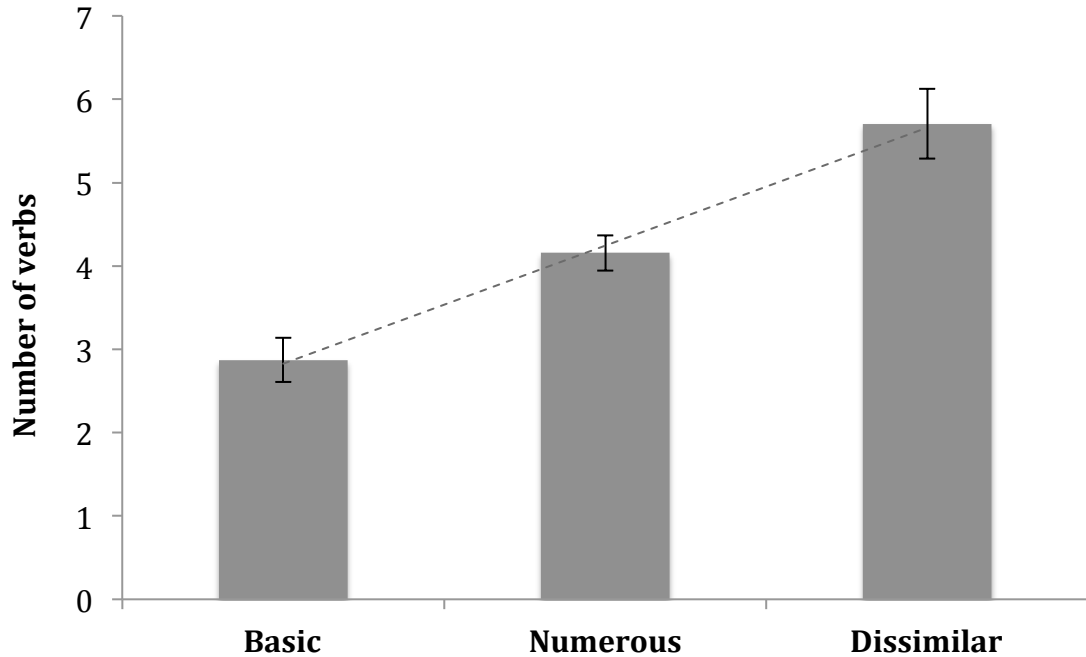


Figure 11. Average number of verbs per item per condition. Error bars represent standard error. Trend line represents linear trend over conditions.

3.5.3 Number of different verbs

The results suggested that the number of different verbs increased over conditions (Figure 12). Results from a repeated measures ANOVA with the number of different verbs aggregated up to the item level as a dependent variable and condition as a repeated factor with three levels (basic, numerous and dissimilar) indicated that there was a main effect of condition ($F(2,28)=20.17$, $p<.001$, $\eta_p^2=.59$). The difference remained significant if the number of different verbs was first divided by the total number of verbs, obtaining a type-token ratio that can be seen as a measure of the similarity between the verbs ($F(2, 28)=7.90$, $p=.002$, $\eta_p^2=.361$). Furthermore, there was a significant linear trend over conditions ($F(1, 14)=29.46$, $p<.001$, $\eta_p^2=.678$).

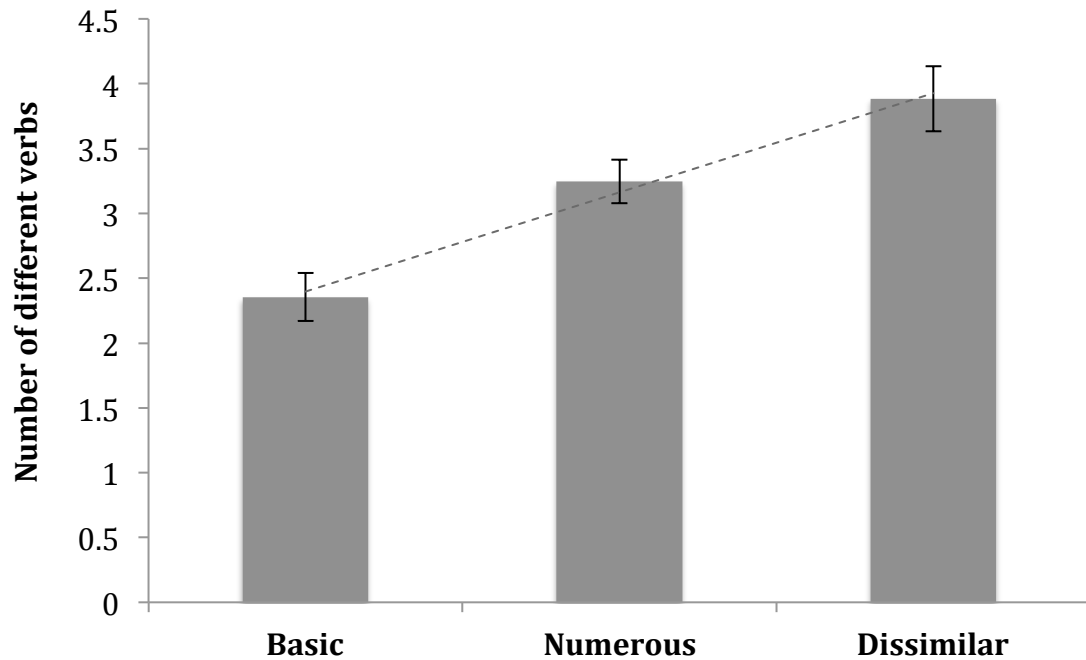


Figure 12. Average number of different verbs per item per condition. Error bars represent standard error. Trend line represents linear trend over conditions.

3.5.4 Correlation analyses

To further investigate the relationship between the linguistic characteristics of the descriptions and the duration ratings, Pearson's correlation coefficient was calculated between the duration ratings and each of the three linguistic measures. Furthermore, the correlations between each of the measures and the segmentation and similarity ratings were calculated to assess whether each measure indeed reflected the assumed event characteristics. None of the measures were related to clock duration, suggesting that the measures reflect animation content rather than clock duration. Results showed that there was a significant relationship between the duration ratings and the total number of words ($r=.35$, $p=.02$). Neither the number of verbs nor the number of different verbs was correlated with the duration ratings. This suggested that the number of words is the main verbal characteristic that is related to estimated duration.

To investigate the relationship between the verbal characteristics and the event properties, correlations between these characteristics and the event

properties were calculated. Results showed that the number of words was also correlated with the number of event boundaries ($r=.42, p=.004$) and the perceived (dis)similarity ($r=-.46, p=.002$). However, surprisingly, there was no significant correlation between the number of verbs and the number of segments. This suggests that the assumption of previous studies that verbs are markers of event boundaries may not be justified, because if verbs were perceived as event markers, one would expect a positive correlation with number of segments.

Both the number of verbs and the number of different verbs showed a significant relationship with the dissimilarity of the events ($r_1=-.45, p=.002$; $r_2=-.51, p<.001$), suggesting that verbs may denote similarity rather than the number of segments. This negative correlation suggests that as more verbs and more different verbs are used, the events that they describe are more dissimilar.

3.6 Discussion

The results presented above suggest that there was an effect of the content of the animations in terms of identified sub-events and similarity between sub-events on the descriptions that people provide during recall. The number of words, the number of verbs and the number of different verbs all showed a significant increase over conditions. Additionally, although the prediction was that only an increase in the number of different verbs would indicate greater dissimilarity, all three measured displayed a relationship with similarity: lower similarity within the event structure is related to more words, more verbs and more different verbs.

Furthermore, only the total number of words displayed a significant positive relationship with the number of segments: more perceived segments lead to more words. Surprisingly, the number of segments was not correlated with the number of verbs or number of different verbs. This suggests that the relationship between the number of sub-events and the number of verbs may not be as straightforward as assumed: more events may not necessarily be denoted by more verbs. The positive relationship with the number of words suggests that sub-event structure may be captured by other words than verbs.

Thus, only the total number of words displayed a relationship with both the number of segments and similarity. This was also the only one of the three measures that correlated significantly with the duration ratings. Together, these findings suggest that there is a relationship between the duration ratings and the number of words used: more words suggest more segments and less similarity between them, and are correlated with longer duration ratings. This is in line with the hypothesis that an increased amount of stored information leads to longer duration estimates, as more words can be seen as a measure of more information (Ornstein, 1969; Pedersen & Wright, 2002). Together, the results from Experiment 1 and 3 suggest that the amount of information encoded about the number of sub-events and the similarity between them affects both the number of words used to describe the remembered events and how long we estimate the duration of the remembered events to be, with more sub-events and less similarity between them leading to more encoded information, both visually and verbally.

These findings raise several questions, namely, whether the total number of words can be matched across conditions and yet convey different underlying event structures, and if so, whether the underlying event structure can still lead to different duration ratings. This is important because it can disentangle the roles of words from memory content, which should in principle, be separable, and thus the amount of information stored in terms of sub-event and similarity structure should be separable from the language. Indeed, the results above suggest that in attempting to provide accurate descriptions of the events, which are not easily described as would familiar events, participants needed more words to indicate the visual characteristics that change over time. However, language allows expressing complex underlying event structure in simple ways, e.g., *building a house*, and there might be similar simple ways to describe our stimulus animations that do not require reference to very detailed visual changes. Therefore, we next investigated these issues and tested whether acceptable descriptions of these animations, which are nevertheless matched for the number of words, lead to differences in memory representation and duration estimation due to underlying event structures, rather than the number of words per se.

3.7 Experiment 4: Duration ratings based on verbal descriptions.

3.7.1 Research hypotheses and aims

The second experiment presented here aimed to identify whether length- and speed-matched descriptions of equally long events with different underlying event structures lead to different duration estimates. This study investigated whether aspects beyond the language used, namely the content of the original animation in terms of number of segments and perceived similarity, predicted duration ratings, over and above the language used. As these underlying event properties have been shown to drive the effects found in Chapter 2, the same properties may also predict duration judgements based on descriptions of these stimuli.

To investigate this, the same experimental paradigm as presented in Experiment 1 was used, but instead of using animations, the present study used descriptions of these animations, accompanied by a still frame from the beginning of the animation. Participants studied the descriptions, performed a memory recognition task and then performed a surprise duration judgement task, basing their duration estimate on their memory representation of what had been described. Therefore, the research questions and hypotheses were similar to the ones in Experiment 1 and 2. The first prediction was that response latencies are affected by the conditions, displaying a main effect of condition and a positive linear trend, as more associated event properties need to be reactivated when the number of sub-events and the dissimilarity between them increases. The second prediction was that there is an effect of condition on reconstructed duration, with more identified event boundaries and less similarity between segments in the original animation leading to longer duration estimates.

Furthermore, this study aimed to identify what aspects of the descriptions predict the duration ratings compared with the results from Experiment 3. As the number of words between conditions was kept equal, the focus was on the number of verbs and number of different verbs, because both showed an association with

similarity in Experiment 3. As before, to identify whether these aspects of the stimuli predict the duration ratings, regression analyses were conducted, investigating the relative amount of variance explained by these characteristics.

3.8 Methods

3.8.1 Participants

Fifty native English-speaking students from the University of York participated for course requirement or a small monetary reward. Three participants were excluded because they had low accuracy scores in the probe recognition task (recognition accuracy $\leq 50\%$ in one of the conditions or a false alarm rate above 50%). Two additional participants with poor memory performance were excluded for the purpose of counterbalancing lists (excluding the two participants with the next worst memory performance from those lists; both with false alarm rates above 40%). Overall correct recognition was 92% (SD=7.3%) after exclusions. These percentages did not differ significantly across item conditions (basic: M=97%, numerous: M=91%, dissimilar: M=87%; Friedman's test n.s.). Participants had normal or corrected-to-normal vision. This experiment was approved by the Ethics Committee of the Department of Psychology of the University of York. Participants provided informed consent and were debriefed after the study.

3.8.2 Materials

The same animations were used as in Experiment 1. Also, the same probes (still frames from the beginning of the animation that are equal between conditions) and foils were used as presented in Experiment 1. The probes were paired with descriptions that were constructed as described below (section 3.8.2.1), instead of with the visual animations.

3.8.2.1 Stimulus creation

Stimulus creation consisted of two steps: (1) obtaining naturalistic descriptions, and (2) finding a representative description for each condition for each item. Firstly, a new group of native English speaking participants (N=24) was recruited online and asked to provide a description for each animation. Participants only saw one condition of each item (counterbalanced through Latin square design). Participants were instructed to describe each animation so that another person who has not seen it can reconstruct what happened when reading the description and viewing the still frame that was provided together with the animation. Participants were told to use a specific number of words, defined by the average number of words used to describe an item across conditions from Experiment 3, plus or minus two words. For each item and each condition, the experimenters picked the first three descriptions that were accurate (and representative; for instance, a description like “Wrecking ball on chain is pulled back and released. The forces of gravity propel it into the centre ball which transfers the immense force to the left ball” was not deemed representative when all other descriptions denoted the same events in terms of circles hitting each other, describing their pattern of motion), within the correct word range and that did not contain any information about the visual display (rather than on the dynamics of the event). These constraints were chosen so as to rule out the possibility that differences between conditions could be due to words being used to describe the display rather than the events (e.g. “There are two stacks of boxes.”). In a few cases, the experimenters had to remove this information to obtain three suitable descriptions. The three descriptions and their accompanying animation were presented to a new group of native English speaking participants (N=28) who were asked to pick the description that they thought best described the animation. As before, participants only saw one condition per item. The preferred description was then used in the main experiment.

3.8.2.2 Stimulus pre-test

Previous studies have suggested a relationship between the number of words and the duration estimate (Experiment 3; Pedersen & Wright, 2002). To verify that the obtained descriptions were equal in terms of number of words, a repeated measures ANOVA was conducted with number of words as a dependent variable. As expected, results showed that there is no significant difference over conditions ($F(2,28)=1.0$, $p=.38$, and all planned comparisons non-significant, $p\geq.23$). Thus, there were no significant differences in number of words between the conditions.

Burt (1999) has shown that descriptions that use 'fast' words (i.e., denoting fast actions) are rated as shorter than descriptions that use 'slow' words. To verify that the obtained descriptions were equal in the speed that they implied, a new group of online recruited native English-speaking participants ($N=30$) was asked to provide speed ratings. Participants only saw one condition of each item. They were given each description and its associated still frame and they were first asked to read each description and visualise the unfolding the events. They were then instructed to rate each of the descriptions on a 1-7 scale (1=very slow, 7=very fast). The aim of this analysis was to investigate the stimulus characteristics, so only the by-item analysis was deemed relevant. A repeated measures ANOVA showed that there was no difference in implied speed between the conditions ($F(2,28)=.70$, $p=.50$, all planned comparisons non-significant, $p\geq.25$). Thus there were no significant differences in implied speed between the conditions.

3.8.3 Design and procedure

The design was largely the same as described for Experiment 1, but instead of studying animations, participants studied still frames (the probes from Experiment 1) accompanied by a description (see Appendix A). To keep the procedure of this experiment as similar to the procedure of Experiment 1 as possible, participants studied the probe on its own for 2 seconds before they saw the description and the still frame together. Stimuli were presented and responses were recorded using E-Prime2 (version 2.0.8.90). Participants were instructed to press a button when

they were done studying the description before moving on to the next frame-description pair. To facilitate memorising the descriptions, after all descriptions in a list were presented, participants were prompted to press a button when they were ready to read the descriptions again. All frame-descriptions pairs were studied three times and each study cycle presented the pairs in random order.

In the recognition memory phase of the experiment, participants were asked to decide whether a presented still-frame belonged to one of the described events that they studied. As in Experiment 1, each participant was presented with 15 probes and 15 foils. Reaction times over 3 standard deviations above the mean for each condition were excluded, as these were considered outliers. Only correct responses were taken into consideration. In the duration rating phase of the experiment, participants were asked to rate the duration of the events as described by the description on a seven-point Likert scale (1 = *very short*, 7 = *very long*). Participants were instructed to use the whole scale. Participants were not aware beforehand of the fact that they had to provide a duration rating until the duration rating phase.

Participants were tested individually in test booths. They were told that they were participating in a memory experiment using descriptions to investigate how we encode memories of events. They were instructed to study the descriptions together with the still frame, imagining the unfolding of the events as described by the language. To ensure that participants studied the descriptions, they were told that they would later be asked questions about their content. They were instructed to study the probe that precedes the description in association with the description, as this would be used in a later part of the experiment to refer back to the description. Participants were debriefed about the actual aims of this study after the experiment.

3.9 Results

3.9.1 Recognition results

The prediction was that response latencies would show a main effect of condition and a positive linear trend. Repeated measures ANOVAs with recognition latencies aggregated up to the subject (F_1) or item (F_2) level as a dependent variable revealed main effects of condition ($F_1(2,90)=10.16, p<.001, \eta_p^2=.184$; ($F_2(2,28)=3.35, p=0.05, \eta_p^2=.193$) (Figure 13). Furthermore, the results displayed a significant linear trend when calculated over subjects, and marginally when calculated over items ($F_1(1,45)=13.95, p=.001, \eta_p^2=.24$; $F_2(1,14)=3.15, p=.10, \eta_p^2=.18$).

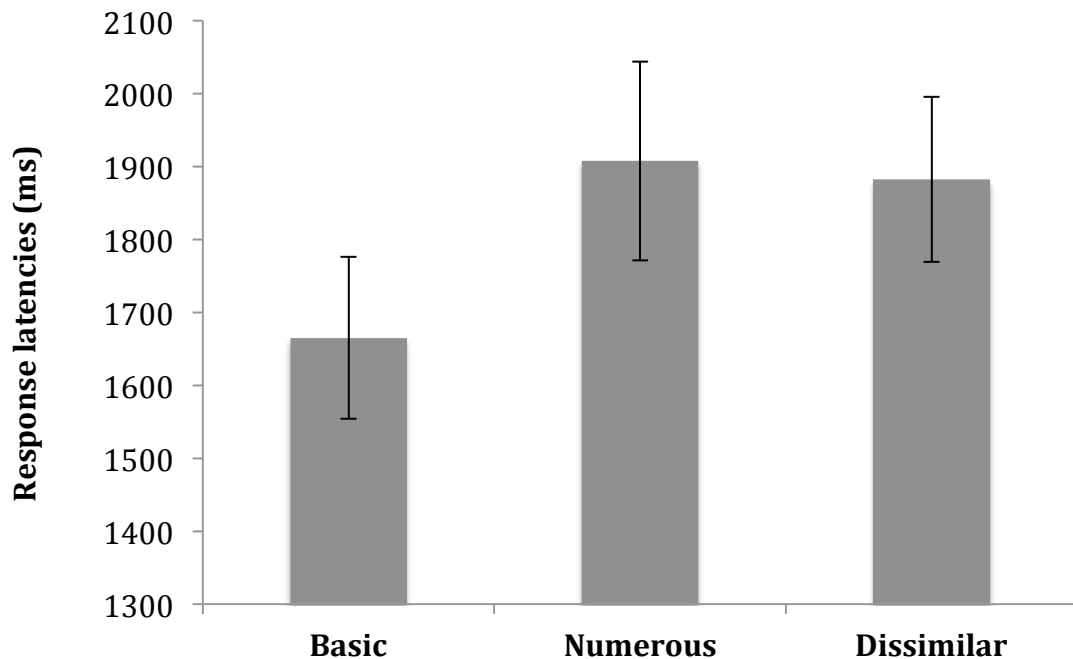


Figure 13. Response latencies over subjects. Error bars represent standard error.

3.9.2 Duration ratings

The prediction was that a main effect of condition and an increasing trend over conditions would be observed for the duration estimates. Repeated measures ANOVAs with duration ratings aggregated up to the subject (F_1) or item (F_2) level as a dependent variable revealed main effects of condition ($F_1(2,88)=9.07, p<.001, \eta_p^2=.17$; $F_2(2,28)=4.60, p=.02, \eta_p^2=.25$) (Figure 14). Furthermore, the results displayed a significant linear trend over conditions ($F_1(1,44)=19.11, p<.001, \eta_p^2=.30$; $F_2(1,14)=6.79, p=.02, \eta_p^2=.33$).

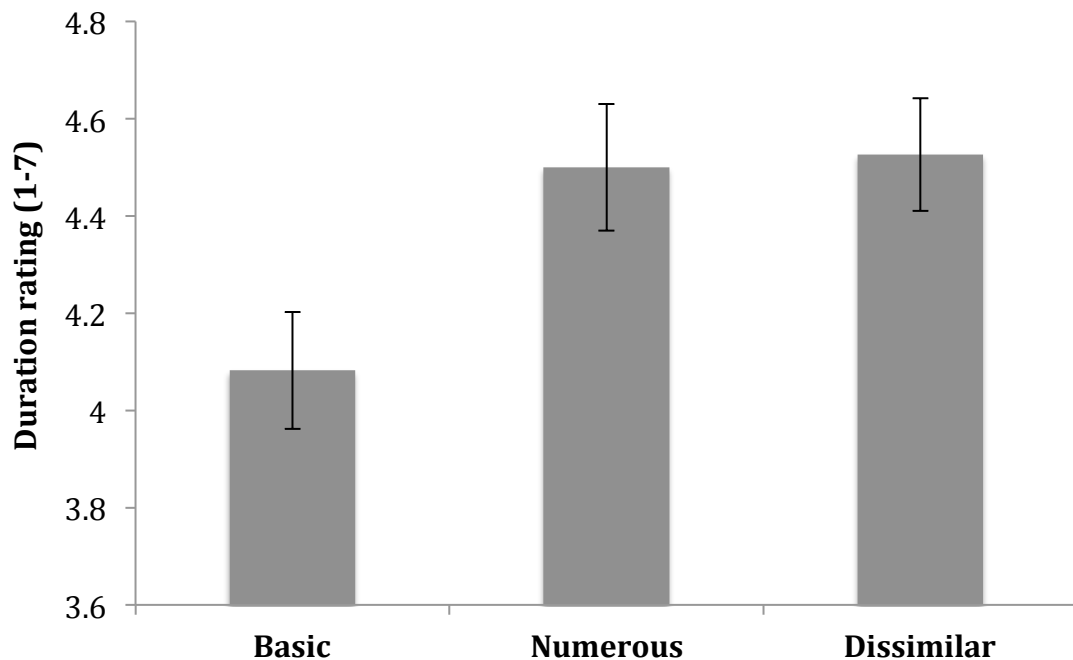


Figure 14. Duration ratings over subjects (scale 1-7). Error bars represent standard error.

3.9.3 Content analyses and regression results

To further investigate the effect of the content of the descriptions on the duration ratings, two measures of content were calculated for each of the description stimuli: the number of verbs and the number of different verbs. Pearson's

correlation coefficient was calculated to investigate the relationship between each of these measures and the duration ratings. There was a marginally significant positive relationship between the duration ratings and the number of different verbs ($r=.28$, $p=.06$). No significant relationship was found between the duration ratings and the number of verbs.

Furthermore, the relationship between the number of identified event boundaries and the similarity between segments that was perceived in the original animations and the duration ratings was investigated. If the descriptions conveyed the underlying event structure of the animations, similar correlations as those found for visual perception would be expected. Indeed, there was a significant relationship between the duration ratings and the number of segments ($r=.61$, $p<.001$). The number of segments was not significantly correlated with any linguistic measure. Surprisingly, there was only a marginally significant relationship between the duration ratings and perceived similarity ($r=-.25$, $p=.09$). Perceived similarity was not correlated with the number of different verbs.

The marginal relationships between the duration ratings and the number of different verbs and similarity ratings suggests that similarity structure (both in terms of verbal description and described visual similarity) might not play a role in estimating duration based on descriptions. To therefore further investigate the effect of the number of different verbs in the descriptions and described similarity structure on duration ratings, two by-item hierarchical multiple regression analyses were conducted. The aim of these analyses was to investigate whether either of these measures explained any variance over and above clock duration. The results showed that neither the number of different verbs (F_1) nor described similarity (F_2) significantly increased the proportion of variance accounted for over and above clock duration ($F_{1\text{change}}(1, 42)= 1.65$, $p=.21$; $F_{2\text{change}}(1, 42)= 2.19$, $p=.15$). As expected given the strong correlation between the number of segments and the duration ratings, adding the number of segments to the regression model did significantly increase the proportion of variance accounted for over and above clock duration from .40 to .67 ($F_{\text{change}}(1, 42)= 21.70$, $p<.001$) (Table 5). This pattern

remained the same if the number of different verbs or the similarity ratings are added to the model first.

These results suggest that contrary to the findings from Experiment 1, similarity does not appear to play a role in estimating duration based on remembered described events. Furthermore, these results suggest that the number of segments plays a role in estimating duration based on a verbal description, over and above the number of different verbs that are used. Thus, despite the fact that the content of the animations is conveyed by words, and that verbs typically refer to actions, the number of different verbs does not seem to explain the duration ratings. The mental representation of the described events goes beyond superficial linguistic properties and thus gives rise to the duration estimates.

Table 5. Regression coefficients for duration ratings. Note: ** indicates $p \leq .01$, * indicates $p < .05$

	Model	<i>B</i>	<i>SE B</i>	β
1.	Constant	3.48	.34	
	Clock duration	.13	.05	.40**
2.	Constant	2.94	.31	
	Clock duration	.09	.04	.27*
	Number of segments	.14	.03	.55**

3.10 Discussion

The results presented here suggest that descriptions of events of equal clock duration but different underlying event structure elicit different duration estimates: events with more segments elicit longer duration estimates. However, there was no effect of similarity: neither the similarity ratings nor the number of

different verbs explained any variance over and above clock duration. This pattern of results suggests that the duration ratings were best explained by the number of segments in the original visual animations, over and above the effect of the use of verbs. Overall, the results suggest that contrary to the visual studies presented in Experiment 1 and 2, there is no effect of perceived similarity between sub-events when participants base their duration ratings on verbal descriptions. The response latencies in the recognition memory task followed the same pattern of results over conditions as the duration estimates, displaying a main effect of condition and a significant difference between the basic and dissimilar conditions, suggesting that some event information was recollected during this task, with more information leading to longer latencies.

3.10.1 Stimulus choice

The present results raise several questions. An important question is whether the results are by-product of the stimulus choice, or whether this is a systematic difference between extracting the temporal unfolding of events from (memories based on) language and vision. Although the present study does not provide conclusive evidence to rule out either possibility, there is evidence to suggest that our stimuli did capture the event structure. Firstly, the experimenters picked three possible descriptions for each condition for each item that were accurate. Furthermore, the most appropriate description out of the three was then picked by a separate group of participants. A post hoc inspection of the stimuli suggests that indeed dissimilarity could have been conveyed by the stimuli in a linguistic manner not captured by any of the measures focused on here. For example, similarity may have been captured through summation of elliptical clauses (e.g. 'it turns blue, then yellow, then red'), which reduces the number of verbs used in general, or by using the same verb to denote dissimilar events ('it turns blue, then it turns yellow and then it turns red'). Furthermore, dissimilar events may have been summarised ('every time this happens, either a or b happens'). Thus, it appears that similarity was not explicitly indicated in any of the linguistic measures that this study focused

on but was nevertheless conveyed, suggesting that building an event representation based on a verbal description may rely on different event properties than one that arises from remembering visual events.

Moreover, this illustrates an important point: crude linguistic measures such as the number of verbs or number of different verbs may not capture event dynamics very well. Even when a more elegant measure such as the number of propositions is used, structures such as ‘every time this happens, either a or b happens’ will not be explicit indicators of more events or more dissimilar events. This has important implications for previous studies as well: counting (action) verbs (cf. Burt, 1999; Pedersen & Wright, 2002) may not be a sensitive measure of the amount of action, potentially attenuating the observed relationship between language and duration estimation. Counting the number of described (sub-)events may be a better measure, given the strong relationship between the number of segments and the duration ratings.

3.10.2 Compression of information in vision and language

If the similarity structure was indeed comprehensively captured by the stimuli, the question remains why in contrast with Experiment 1, this study did not find an effect of similarity on the response latencies and duration ratings. This would imply a systematic difference between reconstructing duration based on linguistic descriptions and visually remembered events. One speculative option is that this is due to differences in how the information is (de)compressed. When visually witnessing a series of events, each of the events has a direct contribution to the percept of the temporal unfolding of the events. This information may then be stored in a more compressed way, guided by the identified event boundaries and similarity between sub-events, with repeated changes allowing for more efficient encoding than dissimilar changes (Eagleman & Pariyadath, 2009; Orbán et al., 2008). As argued in the previous chapter, this compression may be responsible for the observed increase in reconstructed duration. However when reading about a series of events, the temporal unfolding of events needs to be reconstructed from

the language, rather than extracted from visual input. This means that the reader needs to ‘decompress’ the information as conveyed by language in order to obtain information about the temporal unfolding. The economic advantage in the efficiency of encoding the numerous condition over the variable condition may then be diminished: both repeated changes and dissimilar changes need to be reconstructed in order to obtain information about the temporal unfolding of the events. Further research is necessary to address whether this is indeed the case.

3.11 Conclusion

The aims of the two experiments presented above were to explore the relationship between event descriptions and duration estimates, focusing on the relationship between characteristics of descriptions and event properties, and to compare how people estimate duration based on their memory representation of what they have seen and their memory representation based on what they have read. Experiment 3 explored what characteristics of descriptions correlate with duration estimates, and indicated that only the number of words (and not the number of verbs or different verbs) was correlated with duration ratings. Furthermore, the number of words was correlated with event properties such as the number of identified event boundaries and the similarity between them, suggesting that the number of words may be a post-hoc indicator of the amount of information stored about the events.

The finding that the number of verbs and different verbs are not predictive of duration ratings is corroborated by the findings from Experiment 4, which again showed that the relationship between the use of verbs and estimated duration is weak. This finding highlights the limitations of previous studies that investigated number of verbs rather than event properties, suggesting that not all event dynamics are expressed through the use of verbs. Furthermore, Experiment 4 showed that language and vision may differ in how event representations arise and how event information gleaned from exposure contributes to duration estimation: when reconstructing duration based on an event representation that has arisen from a verbal description, there is no effect of the dissimilarity of the sub-events on

the estimated duration. However, duration ratings based on event representations from visual and verbal stimuli both show an effect of the number of identified sub-events, with more events leading to longer attributed durations. This suggests that in both language and vision, the number of perceived events modulates duration estimation.

Chapter 4

Role of event structure in prospective timing

This chapter presents two studies investigating the role of event structure in duration estimation in prospective paradigms. The question is whether and how event properties such as the number of event boundaries and the similarity between them affect duration estimates when participants attend to time. Experiment 5 investigated whether event properties have a similar effect on duration estimates when participants attend to both duration and content while encoding stimulus properties as they have in a retrospective paradigm. Experiment 6 investigated whether there is an effect of content when participants only attend to time, thus explicitly encoding stimulus duration but not (or only implicitly) stimulus content.

4.1 Introduction

The aim of the present chapter was to investigate the role of event structure in prospective paradigms. Traditionally, views on prospective versus retrospective duration estimation have argued for a double dissociation between the two: prospective timing has been thought to critically rely on attention, whereas retrospective duration reconstruction has been thought to be a pure memory process (Block & Zakay, 1997). However, more recently, these views have been challenged. Prospective duration estimation may be affected by episodic memory encoding (Waldum & Sahakyan, 2013). However, it is currently unknown whether firstly, event structure in terms of number of event boundaries and similarity between segments affects prospective duration estimates, and secondly, whether these effects are similar to the effects observed in retrospective timing. Therefore, these questions are addressed in the current chapter. The following section will give an overview of literature on prospective duration estimation, focusing on evidence for shared resources between memory and attention in prospective

timing, and on evidence for a role for episodic memory in the encoding of stimulus duration in prospective paradigms.

4.2 Attending to time

As outlined in Chapter 1, there is a distinction between prospective and retrospective paradigms in time research. Critical to this distinction is whether participants are aware of the fact that they need to monitor time (prospective) or not (retrospective). Traditionally, prospective and retrospective duration estimates have been argued to be affected differentially by task demands. Prospective estimates are thought to rely on attentional processes, whereas retrospective estimates have been argued to solely rely on a memory representation of content. Evidence for this dichotomy comes from studies comparing prospective and retrospective paradigms that have generally shown that attending to time (or time keeping during stimulus presentation) elicits longer estimates than retrospective judgements (Hicks et al., 1976; Block & Zakay, 1997). Further evidence for this dichotomy comes from studies investigating factors that affect duration estimates in both paradigms. For example, a meta-analysis by Block, Hancock and Zakay (2010) has shown that prospective duration estimates are affected by processing difficulty (i.e., attention diverted from time keeping to a secondary task) and not by remembered information, whereas retrospective duration estimates are affected by remembered information (i.e., stimulus complexity) and not processing difficulty.

However, Tobin, Bisson and Grondin (2010) suggested that this dichotomous view on paradigms in time perception research might be oversimplified. They argued that even though most authors would agree that the main distinction between prospective and retrospective paradigms is attentiveness to time, it would be more useful to see these paradigms as the extremes of a continuum of attentiveness to time. For instance, within the area of prospective time research levels of attentiveness to time vary as an effect of interference of the concurrent non-temporal tasks. This suggests that rather than viewing

attentiveness to time as binary (yes/no), it would be better described in terms of “how much” attentiveness to time there is in an experimental paradigm. Based on these observations, Tobin and colleagues concluded that retrospective paradigms should be seen as being on the low end of the spectrum of attentiveness, whereas prospective paradigms vary in the amount of attentiveness to time depending on concurrent tasks.

4.2.1 Attention and cognitive demands in prospective timing

How much attentiveness to time there is also depends on the demands of the situation, for instance to what extent time is relevant to the person who will make the judgment. According to Zakay (1992) temporal relevance is defined by the degree to which taking duration into account is essential for interpreting the meaning of a situation or for decision making (in: Block & Zakay, 2008). Thus, similar to what Tobin and colleagues have argued, the amount of attention that is directed to time is defined by the demands of the task. A much supported idea in time research is that this selectivity in attention is due to the existence of a limited pool of attentional resources, all of which are involved in many aspects of cognitive processing (Kahneman, 1973; Brown, 2008). This suggests that the same attentional resources are used for attending to both temporal and non-temporal cues, implying that when a participant allocates more attention to non-temporal aspects, less attention can be paid to temporal aspects of the stimuli and vice versa.

This also suggests that temporal judgments are susceptible to interference effects (Brown, 1997). When attentional resources are shared between several processes, fewer resources can be allocated to time keeping, resulting in a reduction of the amount of accumulated temporal information. This reduction in the amount of temporal cues then leads to a shorter judgment of duration (Brown & Merchant, 2007). These interference effects become visible when participants perform a concurrent non-temporal task while attending to duration. Studies using dual-task paradigms have shown that the difficulty of a task affects the experienced duration. Participants typically show more variability and error in their duration

estimates or reproductions (Brown & Merchant, 2007), and tasks that are more attentionally demanding elicit shorter estimates of duration (Hicks et al., 1976; Zakay & Block, 1997; Block & Zakay, 2008).

For instance, Sawyer, Meyers and Huser (1994) found that the degree to which participants under-reproduce a temporal interval in a prospective paradigm varies inversely with the difficulty of a concurrent non-temporal task: intervals are significantly more under-reproduced when participants perform a cognitively demanding task such as solving anagrams than when they perform a less demanding task such as identifying letter A's or making X's. Similarly, Sawyer and colleagues (1994) have shown that the same is true for other cognitively demanding tasks such as the Stroop-task versus a black word task. Furthermore, their results indicate that there is no difference in performance between participants who performed no concurrent non-temporal tasks and participants who performed a very simple concurrent task such as producing X's. Thus, the findings by Sawyer and colleagues suggest that attention and cognitive load affect duration judgments, with more diverted attention leading to shorter estimates.

This was corroborated by Brown and Boltz (2002) who compared coherent and incoherent melodies (created by randomly substituting sounds), and coherent and incoherent narratives (created by randomly substituting noun phrases) under conditions where participants had to either perform a concurrent task of different levels of difficulty (error detection or monitoring for one or two types of words) or no concurrent task. Their results suggested that concurrent tasks that require attention to the content of the stimuli elicit shorter duration reproductions. Also, their results indicated that this effect is larger for more difficult secondary tasks, and that performing a simpler concurrent task with low cognitive demands interferes less with the time monitoring task.

In sum, the amount of attention devoted to time keeping depends on whether or not there is a concurrent task, and on how demanding this task is: a more demanding concurrent task detracts more attention from time monitoring and thus leads to shorter duration estimates. Attending to aspects of the stimuli other than time has been shown to detract from time keeping. However,

monitoring the succession and temporal organisation of events and stimulus changes may contribute to the time keeping strategy, as all become more salient to the participant (Brown, 2008). The events that segment time may be stored incidentally (Poynter, 1989). Monitoring the event structure and storing relevant information about it may in that case be a concurrent task that distracts from timing, but at the same time contributes to the time keeping process: event perception itself may guide time keeping by providing temporal markers that can be used as a measure of accumulated time (Poynter, 1983; Liverence & Scholl, 2012). The following section gives an overview of studies looking at the role of memory for event structure and complexity in prospective timing.

4.2.2 Memory and event complexity in prospective timing

Involvement of memory in duration estimation has traditionally been thought to occur exclusively in retrospective paradigms. In a meta-analysis of prospective and retrospective studies, Block, Hancock and Zakay (2010) argued that whether or not participants encode stimulus content “might not affect prospective duration judgements much [...], although [...] it would affect the memory processes that allegedly underlie retrospective duration judgements” (Block et al., 2010, p. 333). For example, using patterns of light flashes, Poynter and Homa (1983) did not find an effect of whether participants were required to remember stimulus content or not on prospective duration estimates.

However, there is some evidence that stimulus content itself does affect prospective duration estimates. In the same study, Poynter and Homa (1983) found that the number of light flashes affect duration estimates: more light flashes lead to longer duration estimates. In terms of stimulus complexity, Zakay (1993) found that when participants rate the duration of tactilely presented stimuli, duration estimates are affected by the complexity of the stimuli, although the magnitude and direction of the relationship depends on the actual duration and type of duration judgement. Data did not allow for a univocal conclusion about this relationship. Similarly, Poynter and Homa (1983) also found that pattern

complexity had an effect on duration estimates, and showed that the effect was non-monotonous over different durations within the task: in some cases (e.g. for shorter durations) this relationship was positive, while for longer duration this relationship was U-shaped. Hence, there appears to be little consensus on what the effect of complexity is on prospective judgements.

Prospective estimates may thus be affected by the number of stimuli and the relative complexity of stimuli, although evidence is not univocal in explaining the nature of this relationship. These observations motivate further research into whether event structure affects prospective duration estimates. Although this area is not widely studied, the relationship between event boundaries and duration perception has been further examined by Liverence and Scholl (2012) who investigated whether subjective time depends on how the experience of time is segmented into events. In their study, participants viewed short animations of animated dots that followed a path, and while moving, they disappeared and reappeared as the same or new objects. Each time participants had watched a short sequence, they were asked to provide a duration reproduction. Hence, the paradigm is prospective in nature, as participants are aware of the duration reproduction task. Their findings indicated that the animations that contained more discrete events were reproduced as being shorter than the ones with one continuous event. This contrasts with the results of Poynter and Homa (1983), who found longer estimates for more stimulus segments.

These conflicting results may be due to differences in tasks and measures used across experiments. Duration reproductions have been argued to inversely correlate with duration estimates in seconds (Carlson & Feinberg, 1968, 1970; Hansen & Trope, 2013), because they are claimed to ultimately depend on whether time is perceived as passing slowly or quickly during the trial. Thus if participants perceive time as passing quickly, they reproduce long intervals, but assign shorter duration estimates, and vice versa. Although the relation between these measures and the speed of time passage remains to be studied in more detail, it is possible that different measures rely in different kinds of representations. It is possible for example that immediate duration reproduction (like in Liverence and Scholl's task)

involves mentally replaying the rate of stimulus presentation or the trial tempo stored in working memory in prospective paradigms. If so, more segments will lead to a shorter reproductions (quicker rate of reproduction) than fewer segments (slow reproduction). In contrast, an (absolute) estimate (in seconds or on a scale) may rely on a representation of the overall memory content of the interval, like in retrospective judgements, and may thus lead to similar findings to those reported in previous chapters (more segments leading to higher estimated duration).

This possibility is consistent with recent evidence suggesting a role for episodic memory in prospective judgements similar to that found in retrospective paradigms. In a study by Waldum and Sahakyan (2013) participants performed a prospective duration estimation task while performing a lexical decision task with music in the background. Their results showed that when participants heard more songs during duration encoding, their prospective estimates were longer. Waldum and Sahakyan argued that boundaries between songs could then be seen as ‘critical boundaries’ (or, in terms of the present thesis, event boundaries), suggesting that memory for discrete events may affect prospective duration estimates. This study thus indicates that when stimulus characteristics are relevant to the timing task, participants may use them to estimate duration.

In sum, results from previous research are in agreement that event structure affects prospective duration estimates, but the exact nature of this relationship remains unclear and evidence is conflicting. Here, the notion that event boundaries may play a role in prospective paradigms was further investigated, exploring the role of event boundaries as temporal markers by investigating whether event boundaries and similarity between sub-events affect prospective duration ratings, and whether implicit and incidental encoding of events elicit the same effects.

4.3 Experiment 5: Remembering event structure in prospective timing

4.3.1 Research hypotheses and aims

The aim of this study was to identify whether event structure, and in particular the number of identified event boundaries and similarity between sub-events, affects prospective duration estimation. In the present experiment, the effect of explicitly encoded event structure was investigated: participants were instructed to encode both the content of the animations for a subsequent memory task, as well as the relative duration of each animation for a duration rating task.

Based on the literature on effects of event structure in prospective paradigms discussed above, an effect of number of events was expected: more events should lead to longer duration estimates (Waldum & Sahakyan, 2013). Furthermore, based on the findings from Experiment 1 and 2, an effect of similarity between sub-events was expected because people are instructed to encode content as well as to judge duration, although previous studies are not in agreement about the direction of this prediction in prospective paradigms. Thus, an effect of event structure was predicted. If event structure affects prospective duration estimation in the same way that it affects retrospective duration estimation, then a pattern of results similar to the results presented in Chapter 2 would be expected, suggesting that participants also employ their memory representation of the underlying event structure when providing prospective duration judgements, and not just when providing retrospective ones. In that case, a main effect of condition and a positive linear trend across conditions would be expected, and the ratings of stimulus properties would be expected to explain a significant amount of variance in the ratings: as the number of perceived segments increases, and as the perceived similarity decreases, duration estimates should increase, as more segments and more dissimilarity between them should lead to more stored information. Furthermore, a similar pattern of results was expected for the response latencies in the recognition task, as these latencies have been argued to—to some extent—reflect

the amount of encoded event information that needs to be recollected when participants are presented with a probe. Note however that this study on its own cannot distinguish between effects of event structure on time keeping and effects of remembering event structure. This will be further addressed in the second experiment of this chapter.

In the current prospective paradigm, participants were instructed to attend to the duration of the stimuli as well as to remember their content. This allowed for comparing the present prospective results with the retrospective results of Chapter 2. Based on the previous studies described above, when more attention is directed to time keeping, the duration ratings would be expected to be higher than when less or no attention is devoted to it (Brown & Boltz, 2002), as shown by dual-task prospective paradigms leading to shorter estimates compared to single tasks and by previous comparisons between prospective and retrospective judgments. Therefore, the prediction was that duration ratings over items for the present paradigm would be higher than the ratings obtained in Experiment 2, Chapter 2.

4.4 Methods

4.4.1 Participants

Eighty-two native English speaking students from the University of York participated for course credit, course requirement or a small monetary reward. Seven participants with poor memory accuracy were excluded as they had low accuracy scores in the recognition task (recognition accuracy $\leq 50\%$ in one of the conditions or a false alarm rate above 50%). The overall recognition memory accuracy was 87% (SD=8.8%) after participant exclusions, taking correct identification into account. There was no difference between conditions as per design (Friedman's test n.s., basic: M=88%, numerous: M=88%, dissimilar: M=87%). Participants had normal or corrected-to-normal vision. This experiment was approved by the Ethics Committee of the Department of Psychology of the

University of York. Participants provided informed consent and were debriefed after the study.

4.4.2 Materials

The same materials were used as in Experiment 2, Chapter 2. Thus, the description of the materials can be found in section 2.7.2.

4.4.3 Design and procedure

The same experimental design was used as in Experiment 2, Chapter 2. The description of the design can be found in section 2.7.4. However, the procedure of the current experiment was different. Participants were recruited to take part in an experiment on time and memory. They were instructed to pay careful attention to the content of the animations and remember what happens in them, and to monitor how long they are relative to each other. Participants were instructed to attend to the relative duration of the animations rather than their real-time duration, and participants were instructed not to tap or count in their head as a way of timekeeping while studying the animations. Hence, participants were explicitly instructed to attend to both the duration and content of the stimuli. Furthermore, participants were aware of the fact that they would later on perform both a memory task as well as a duration judgement task. From the memory recognition task, response latencies of correctly identified YES-responses were analysed and latencies larger than 2.5 standard deviations of the condition mean were removed to control for outliers.

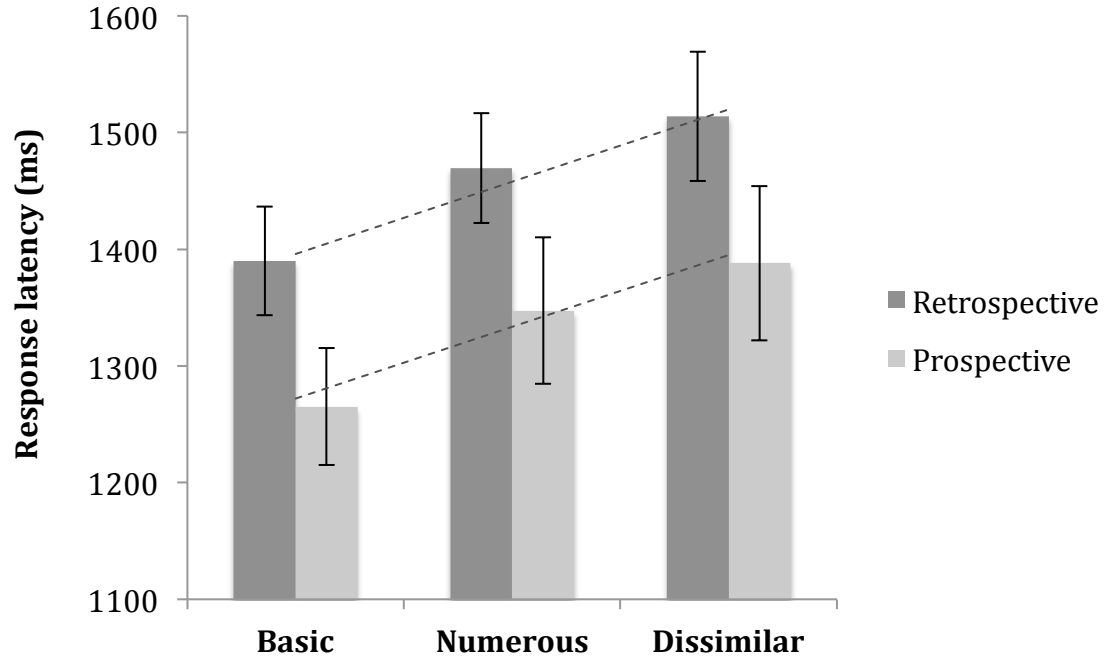


Figure 15. Response latencies over subjects. Error bars represent standard error. Trend line represents linear trend over conditions. Darker colour represents results from Experiment 2 (retrospective), the lighter colour represents results from the present experiment (prospective).

4.5 Results

4.5.1 Recognition memory

The results of repeated measures ANOVAs with response latencies aggregated up to the subject (F_1) or item (F_2) level as a dependent variable revealed main effects of condition ($F_1(1.769, 130.9)=3.628, p=.03, \eta_p^2=.047$ (Greenhouse-Geisser); $F_2(2, 54)=3.58, p=.035, \eta_p^2=.117$) and significant linear trends ($F_1(1, 74)=5.27, p=.025, \eta_p^2=.066, F_2(1, 27)=8.28, p=.008, \eta_p^2=.235$) (Figure 15). These findings are consistent with the patterns found in the previous chapters, as expected, as this pattern reflects the association between the content of the animation and its probe (paired-associate memory cf. Yonelinas 2001), suggesting that properties of the events encoded in memory during the study phase affect memory judgements, with

the conditions that require recollection of more sub-events and less similarity between them to lead to increased response latencies.

Interestingly, as Figure 15 illustrates, the reaction times in the present prospective paradigm were faster than those for the retrospective paradigm presented in Experiment 2, Chapter 2. To investigate this further, a repeated measures ANOVA with response latencies aggregated up to the item level as the dependent variable was conducted and revealed that response latencies for the dual-task learning paradigm ($M=1339.8$ ms, $SE=24.6$ ms) were indeed significantly shorter than those for the retrospective paradigm ($M=1442.1$ ms, $SE=28.6$ ms) ($F(1,81) = 22.52$, $p < .001$, $\eta_p^2=.218$). A repeated measures analysis with data aggregated up to the subject level was not deemed meaningful, as response latencies between experiments were collected from different participant samples.

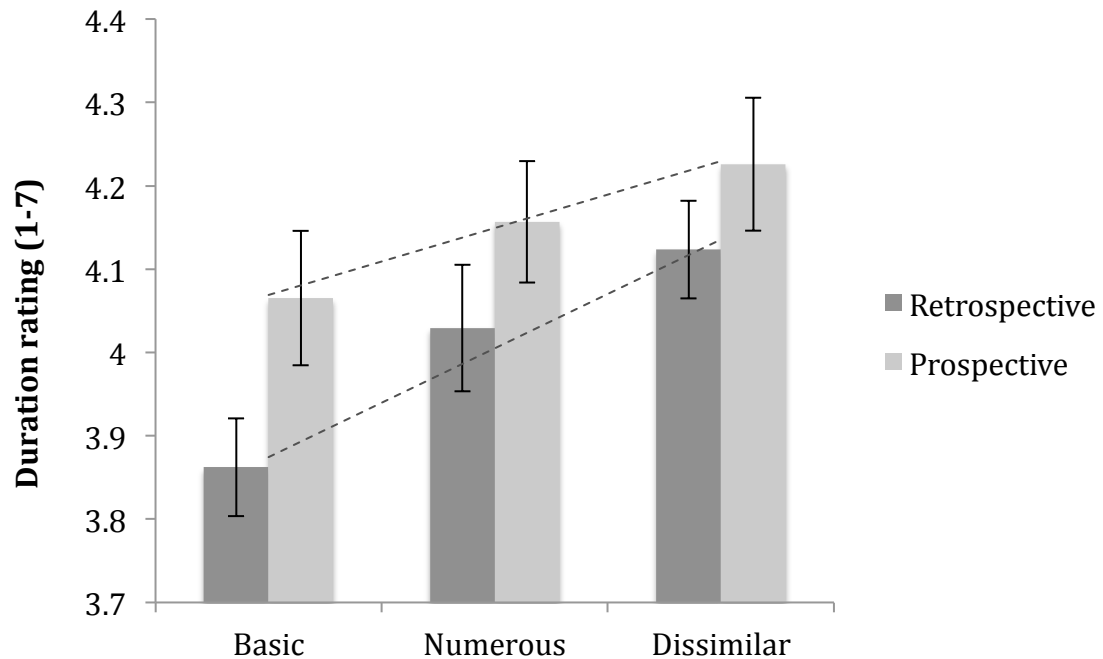


Figure 16. Duration ratings over subjects (scale 1-7). Error bars represent standard error. Trend line represents linear trend over conditions. Darker colour represents results from Experiment 2 (retrospective), the lighter colour represents results from the present experiment (prospective).

4.5.2 Duration ratings between conditions

Repeated measures ANOVAs were used to investigate the effect of condition on the duration ratings. The results of repeated measures ANOVAs with duration ratings aggregated up to the subject (F_1) or item (F_2) level as a dependent variable indicated that there is a main effect of condition on the duration ratings of the animations, although marginally across items ($F_1(2, 148) = 3.65, p = .028, \eta_p^2 = .047$; $F_2(2, 54) = 2.83, p = .068, \eta_p^2 = .095$). Furthermore, the results indicated significant linear trends ($F_1(1, 74) = 6.59, p = .012, \eta_p^2 = .082$; $F_2(1, 27) = 5.60, p = .025, \eta_p^2 = .17$) (Figure 16). These findings were similar to the results from the previous chapter, suggesting that the event properties encoded in memory are retrieved during the duration judgement task.

4.5.3 Duration ratings between paradigms

A repeated measures ANOVA was used to investigate the effect of paradigm on the duration ratings. The results of a repeated measures ANOVA with duration ratings aggregated up to the item level as a dependent variable indicated that there was a significant main effect of paradigm on the duration ratings ($F(1, 81) = 14.53, p < .001, \eta_p^2 = .15$, with pair-wise comparisons indicating that the duration ratings for the current prospective paradigm ($M = 4.15, SE = .09$) were significantly higher than for the retrospective paradigm presented in Experiment 2 ($M = 3.99, SE = .10; p < .001$) (Figure 16). This was in line with the prediction that duration ratings should be higher when more attention is paid to time. A repeated measures analysis with data aggregated up to the subject level was not deemed meaningful, as response latencies between experiments were collected from different participant samples.

4.5.4 Regression results

To further investigate the effect of the event properties on the duration judgements, a by-item hierarchical multiple regression analysis was conducted, investigating the proportion of variance accounted for by the similarity between

events and the number of segments (as obtained in Chapter 2, section 2.7.3) over and above the actual clock duration of the animations. Clock duration was therefore added as a control predictor in the model, as it accounts for the systematic variation between the triads. The results showed that when adding the number of segments to the regression model, the proportion of variance accounted for (R) increased significantly from .63 to .71 ($F_{\text{change}}(1, 81) = 17.38, p < .001$). Furthermore, when the similarity between events was added to the latter model, the proportion of variance accounted for increased significantly from .71 to .73 ($F_{\text{change}}(1, 80) = 5.03, p = .03$) (Table 6). Changing the order in which number of segments and similarity between events were added to the model did not affect the pattern of significance. These findings showed that similar to retrospective duration estimates, duration ratings were modulated by event properties in a prospective paradigm in which participants attended to both time and content.

Table 6. Regression coefficients for duration ratings. Note: ** indicates $p \leq .001$, * indicates $p < .05$.

	Model	B	$SE\ B$	β
1.	Constant	2.54	.23	
	Clock duration	.27	.04	.63**
2.	Constant	2.35	.22	
	Clock duration	.14	.05	.34**
	Number of sub-events	.20	.05	.44**
3.	Constant	2.97	.35	
	Clock duration	.13	.04	.31**
	Number of sub-events	.20	.05	.44**
	Similarity	-.12	.05	-.17*

4.6 Discussion

Similar to the results presented in Chapter 2, the results of the experiment presented here suggest that there is an effect of event structure on duration estimates as illustrated by the increase in duration ratings, and on the recollected event information during probe recognition as illustrated by the increase in response latencies in the recognition memory task. Furthermore, a similar positive linear trend was found over conditions for both tasks. The findings from the regression analyses suggested that the number of encoded sub-events and similarity structure for events of the same clock duration modulate duration estimates, as more sub-events and less similarity between them lead to longer duration judgements. Comparisons across experiments showed that duration estimates are longer when people attend to both time and content, suggesting that more remembered information about stimulus content and its duration leads to longer attributed duration. Furthermore, recognition decisions were faster when participants have attended to both time and content while studying the stimuli.

4.6.1 Comparison between paradigms

A comparison between the retrospective paradigm used in Experiment 2 and the prospective paradigm presented here showed that participants were faster to recognise the correct probes in the present paradigm. This could be due to the fact that the more we know about a stimulus, the easier it is to recognise it, as we are more familiar with it. For instance, familiar faces and words are easier to recognise than their less familiar counterparts (Yonelinas, 2002). Relative to the retrospective paradigm, participants have encoded more information about the stimuli: not only did they encode the content of the animations (as before), they have also encoded additional information about their relative duration, and perhaps more information about the event structure, as the event structure may to some extent guide duration encoding and may thus be more salient (this will be discussed in more detail below). More associated information with an item may then lead to faster recognition times.

However, as argued in section 2.9.1, our task is not simply a recognition task in terms of familiarity. It requires participants to recollect the content of the animation in more detail, as they have to indicate whether a presented still frame was or was not part of the studied animation. This recollection component could explain why the results still display an effect of event structure, with more complex event structures leading to longer reaction times, but relative to the previous study, reaction times are generally faster, because more information has been stored in association with the stimulus during learning, leading to greater familiarity (cf. Yonelinas, 2002).

As hypothesised, duration ratings were higher for the present prospective paradigm than for the retrospective paradigm in Experiment 2: encoding more information, including information about content and about duration, leads to longer duration estimates. This is in line with findings from previous studies showing longer estimates or reproductions when less attention is diverted from timing the stimulus (Brown & Boltz, 2002). However, these data do not allow for firm conclusions about the role of event structure in this prospective paradigm due to the dual task instructions: although the increase in response latencies suggests that there is an effect of event structure on the memory representation of the stimulus, this does not directly imply that prospective duration estimates are generally affected by the memory representation. Instead, in purely prospective paradigms they may be affected by the interplay between the event structure and the time keeping process: when people attend to time, the event structure may become more salient, as events are markers of temporal development. Therefore, event structure may guide time keeping directly, rather than via a memory representation. The following experiment attempted to tease apart these processes by investigating whether event structure plays a role when participants do not memorise content, minimising the role of memory for event structure.

4.7 Experiment 6: The role of event structure in prospective timing

The previous experiment has shown that the underlying structure of events plays a role when people perform a duration estimation task in a paradigm where they pay attention to time while remembering the content of the stimulus in terms of its event structure, showing a similar pattern of duration ratings as the retrospective experiment in Chapter 2. This could suggest that the memory representation of the events affects prospective duration estimation. However, as pointed out above, these findings alone are not sufficient to claim that this effect is due to retrieving event structure from memory rather than an effect of event structure on encoding time, as recent studies have suggested that event structure also appears to modulate time keeping in prospective studies (Liverence & Scholl, 2012; Waldum & Sahakyan, 2013).

This leads to the question whether event structure has the same effect as observed in our retrospective studies and the prospective study above when participants do not pay attention to content. Attentiveness to time is then maximal, while the resources allocated to remembering content in terms of event structure are minimal. The present study therefore investigated whether the number of perceived event boundaries and the similarity between them play a role in prospective duration estimation when participants perform a timing task without being instructed to memorise the stimuli, teasing apart the effect of event structure on time keeping and the effect of encoded of event structure and subsequent retrieval from memory on duration estimates.

4.7.1 Research hypotheses and aims

The main aim of this experiment was to investigate whether event structure plays a role in a prospective paradigm where people only attend to time and not to content. If the results do not show an effect of event structure, this suggests that in a single-task prospective paradigm, participants solely base their estimate on

information obtained by a timing mechanism or strategy. The observations of an effect of event structure on duration estimates in the previous studies could then be attributed to encoding and retrieving detailed representations of the content of events in and from memory.

However, if there is interplay between timing mechanisms and event structure (event boundaries could for example contribute to the timing process by providing temporal markers), it is expected that event structure does play a role. The question then is whether both the number of identified event boundaries and the similarity between them play a role. As outlined above, the number of identified boundaries could affect prospective duration estimates: when participants attend to time, event boundaries become more salient (as they are markers of temporal development) and could therefore contribute to the accumulation of 'time'. Given the conflicting evidence about the direction of an effect of stimulus complexity on prospective duration estimates, it is unclear what the role of similarity is in prospective timing. The hypothesis therefore was that both the number of identified event boundaries and the similarity between sub-events modulate duration estimates.

Furthermore, it was expected that the response latencies in the recognition memory task would not be affected by the conditions. As the effect of condition on the response latencies is thought to arise from the amount of information that is recollected about the content of the probe's corresponding animation, we did not expect to see an effect of condition if content (in terms of event structure) is not encoded. Furthermore, response latencies were expected to be slower than in the dual-task paradigm above, as less information (i.e. only duration, not content) is encoded about the stimuli.

A second question in this experiment was whether duration ratings are affected by the paradigm. Compared with the previous experiment, the present paradigm does not require the participant to attend to both time and content, but only requires the participant to focus on the duration of each animation. Hence, in terms of attentiveness to time, this task should allow a participant to focus more on time keeping than the previous task. Previous attention research (section 4.2.1)

suggested that when there is a secondary task to attend to, resources are shared between the tasks and less attention is paid to tracking duration. Therefore, duration judgements in experiments with a secondary task tend to be shorter than duration judgements in single-task experiments (e.g. Block & Zakay, 1997). It was therefore hypothesised that the duration ratings in this experiment would be higher than those obtained in Experiment 5.

4.8 Methods

4.8.1 Participants

Seventy-nine native English speaking students from the University of York participated for course credit, course requirement or a small monetary reward. Four participants with poor memory accuracy were excluded as they had low accuracy scores in the recognition task (recognition accuracy $\leq 50\%$ overall correct recognition or $> 50\%$ false alarm). Note that this is slightly more lenient than in Experiment 2 and 5 (see section 4.10.2 for further discussion). The overall recognition memory accuracy was 83% (SD=15%) correct taking correct identification into account, after participant exclusions. There was no difference between conditions as per design (Friedman's test n.s., basic: M=81%, numerous: M=85%, dissimilar: M=84%). Participants had normal or corrected-to-normal vision. This experiment was approved by the Ethics Committee of the Department of Psychology of the University of York. Participants provided informed consent and were debriefed after the study.

4.8.2 Materials

The same materials were used as Experiments 2 and 5. The description of the materials can be found in section 2.7.2.

4.8.3 Design and procedure

The same experimental design was used as in the previous experiment. The description of the design can be found in section 2.7.4. However, the procedure of the present experiment was different. Participants were recruited to take part in an experiment on time. They were instructed to pay careful attention to the duration of the animations. Participants were instructed to attend to the relative duration of the animations rather than their real-time duration, and participants were instructed not to tap or count in their head as a way of timekeeping while studying the animations. Importantly, participants were not instructed to remember the content of the animations and were unaware of the subsequent memory test. Hence, participants were explicitly instructed to attend to temporal information but not to the content of the stimuli and they were aware of the fact that they would later on perform a duration judgement task. It was pointed out to the participants that each animation would be preceded by a still frame that would later on be used to refer back to the animation. From the recognition memory data, response latencies of correct responses were analysed (2.5 SD of the condition mean were removed) to investigate whether the response latencies are affected by the conditions, reflecting an effect of paired-associate memory between the probe and content. For the duration estimation task, items where it took participants longer than 10 seconds to respond to provide a duration rating were excluded.

4.9 Results

4.9.1 Recognition memory

The results of repeated measures ANOVAs with response latencies aggregated up to the subject (F_1) or item (F_2) level as a dependent variable indicated that there was no significant main effect of condition on the response latencies ($F_1(2, 148) = .47, p = .63, \eta_p^2 = .006$; $F_2(2, 54) = .26, p = .78, \eta_p^2 = .009$), no significant linear trends ($F_1(1, 74) = 0.07, p = .94, \eta_p^2 = .000$, $F_2(1, 27) = .061, p = .81, \eta_p^2 = .002$) and no significant

contrasts (all p 's $\geq .35$) (Figure 17). These findings are consistent with the prediction there would not be a strong association between the probes and the content of the animations, as content was not explicitly studied here. Thus, little if any content information in terms of event structure was recollected during the recognition memory task.

To investigate the effect of paradigm, a repeated measures ANOVA with response latencies aggregated up to the item level as the dependent variable was conducted and revealed that response latencies for the single task prospective paradigm ($M=1366.7$ ms, $SE=24.6$ ms) were not significantly different from those for the dual-task prospective paradigm above ($M=1339.8$ ms, $SE=24.6$ ms) ($F(1,81)=1.37$, $p=.25$, $\eta_p^2=.017$). However, response latencies for the single task prospective paradigm were significantly shorter than those for the retrospective paradigm ($M=1442.1$ ms, $SE=28.6$ ms) ($F(1,81)=7.44$, $p=.008$, $\eta_p^2=.084$). Repeated measures analyses with data aggregated up to the subject level were not deemed meaningful, as response latencies between experiments were collected from different participant samples.

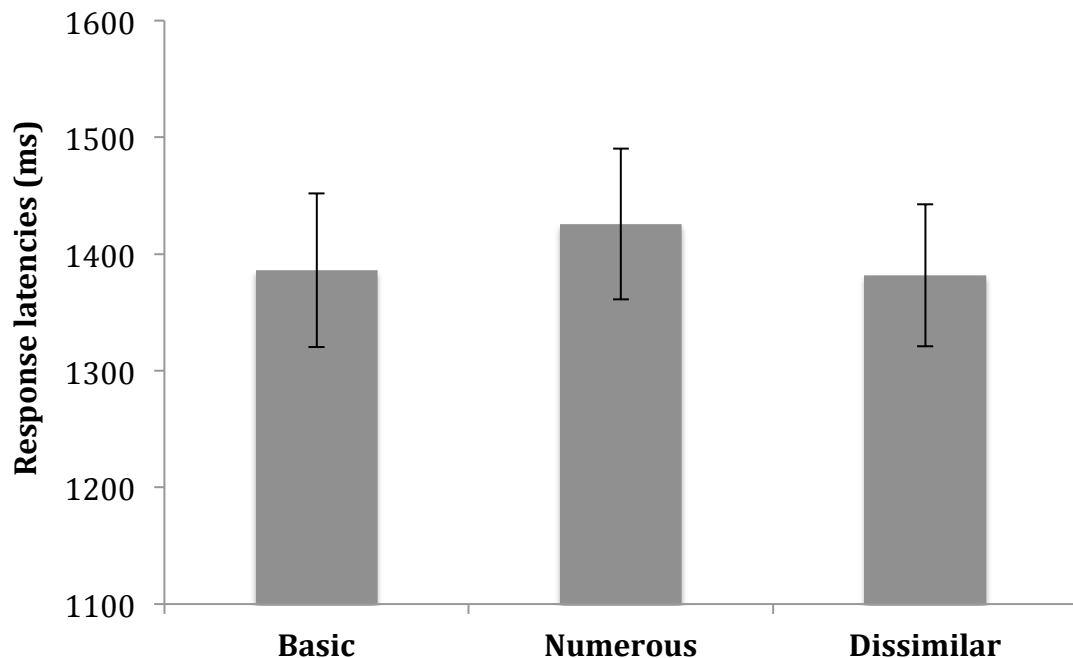


Figure 17. Response latencies over subjects. Error bars represent standard error.

4.9.2 Duration ratings between conditions

Repeated measures ANOVAs were used to investigate the effect of condition on the duration ratings. The results of repeated measures ANOVAs with duration ratings aggregated up to the subject (F_1) or item (F_2) level as a dependent variable indicated that there was no main effect of condition on the duration ratings ($F_1(2, 148) = .51, p = .60, \eta_p^2 = .007$; $F_2(1.66, 44.72^*) = .60, p = .52, \eta_p^2 = .022$ *Greenhouse-Geisser corrected for sphericity). Furthermore, the results indicated no significant linear trends ($F_1(1, 74) = 1.01, p = .32, \eta_p^2 = .013$; $F_2(1, 27) = .82, p = .37, \eta_p^2 = .029$) and no significant contrasts (all p 's $\geq .31$) (Figure 18).

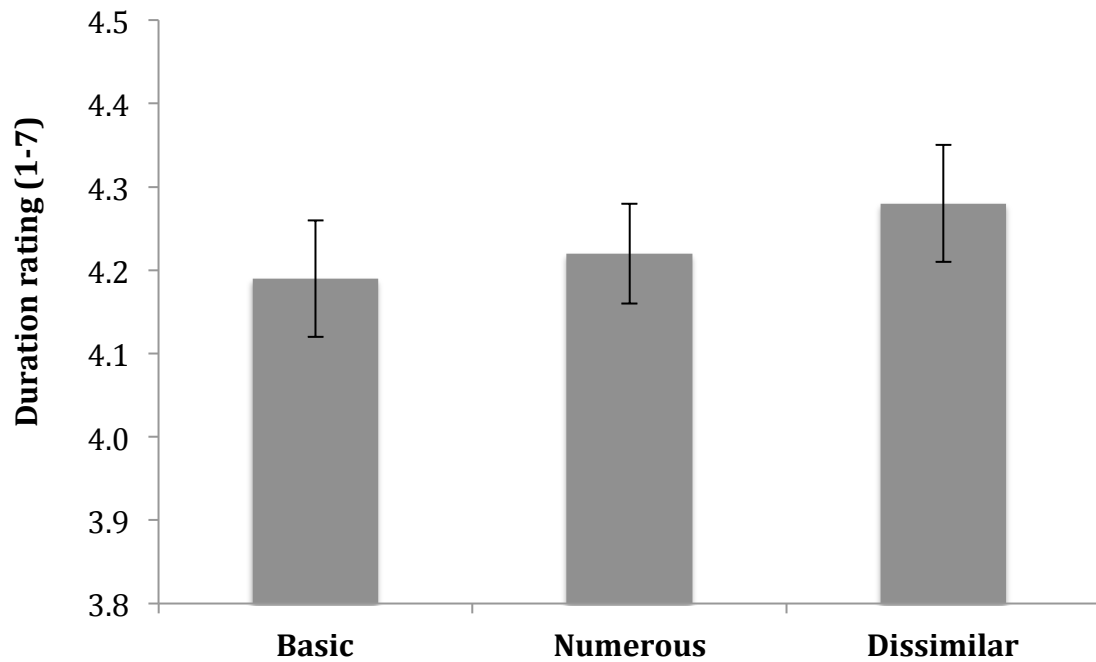


Figure 18. Duration ratings over subjects (scale 1-7). Error bars represent standard error.

However, given that participants may have stored more veridical information about the actual duration of the animations, they may be more comfortable using the whole range of the scale compared to the previous studies, leading to more variance in the data. To verify that this null result is not due to this potential extra variance, ratios between the ratings and the actual durations of the animations

were calculated. Using these ratios over items as a dependent measure, still no differences between conditions were found (main effect, linear trend and contrast all $p \geq .45$) (Figure 19). These findings suggested that in a prospective paradigm where participants only attended to time, duration estimates were not modulated by conditions.

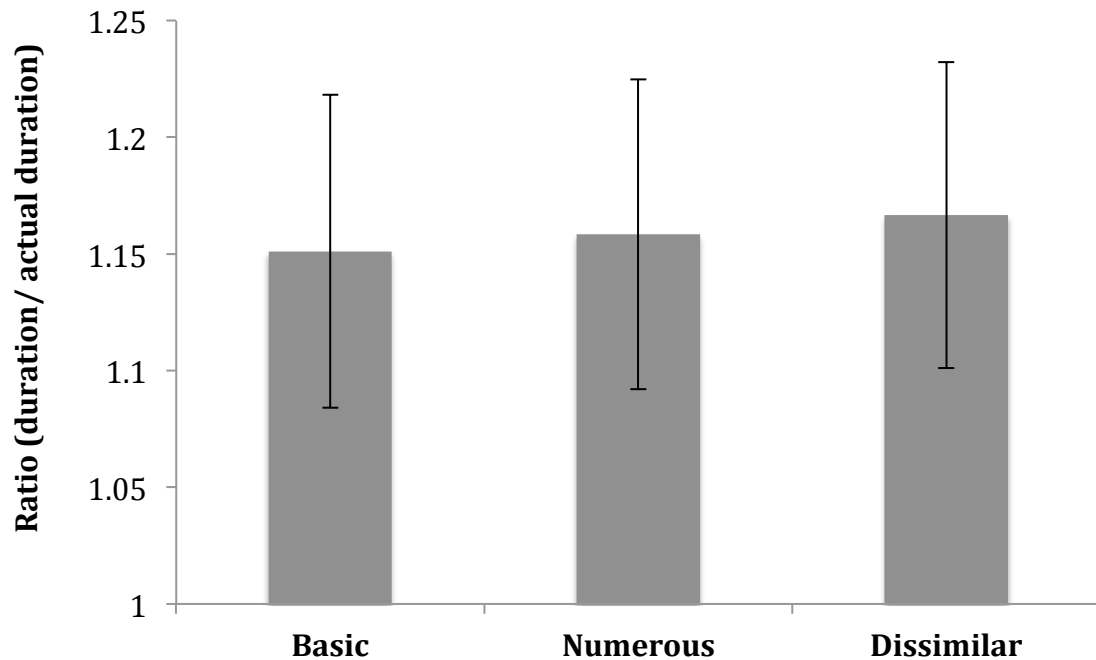


Figure 19. Ratios between estimated duration and actual duration over items. Error bars represent standard error.

4.9.3 Duration ratings between experiments

To investigate whether duration ratings are higher in a single-task versus a dual-task paradigm, we compared the duration ratings over items between this study and the dual-task study described above. Repeated measures analyses with data aggregated up to the subject level were not deemed meaningful, as response latencies between experiments were collected from different participant samples. Results from a repeated measures ANOVA with duration ratings aggregated up to the item level as the dependent variable revealed that there was no significant main effect of experiment ($F(1,81)=1.95$, $p=.17$, $\eta_p^2=.023$). This suggests

that when all three conditions are taken into account, participants do not provide significantly higher ratings in the single task experiment ($M=4.21$, $SE=.09$) than in the dual task experiment ($M=4.15$, $SE=.09$). However, visual inspection of the data suggested that the difference between the two paradigms might be larger for the basic condition than both other conditions (Figure 20). A paired-samples t-test suggested that there might indeed be a trend, with the basic condition being rated as longer in the single task experiment ($M=4.16$, $SE=.16$) than in the dual task experiment ($M=4.04$, $SE=.17$) ($t(27)=1.77$, $p=.09$).

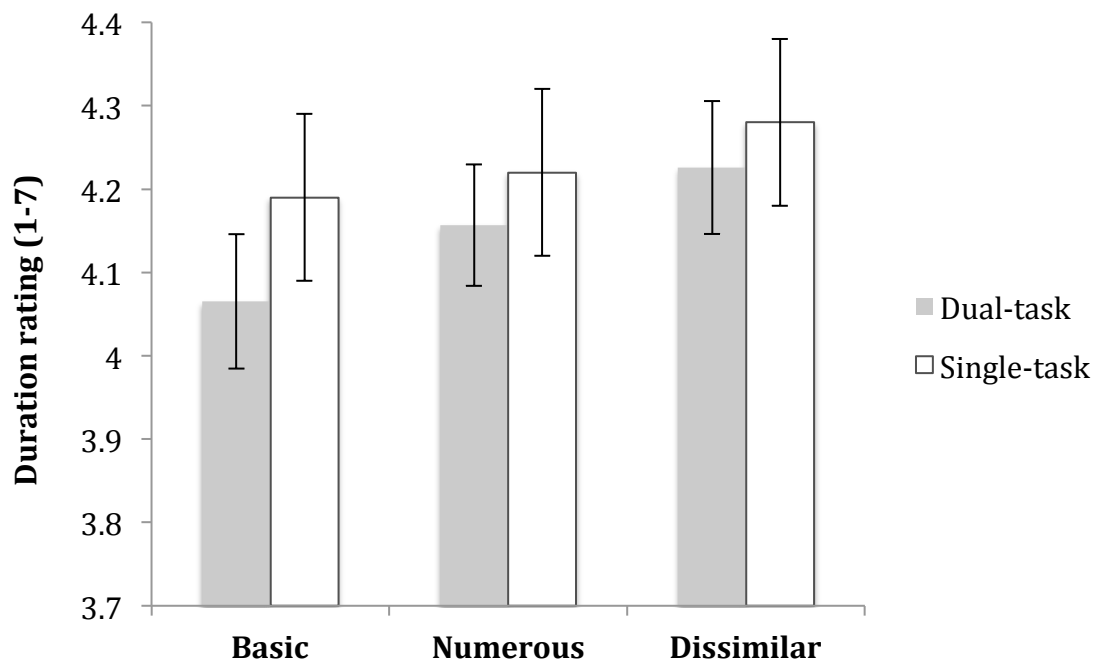


Figure 20. Duration ratings over subjects (scale 1-7). Error bars represent standard error. Trend line represents linear trend over conditions. Darker colour represents results from Experiment 5 (dual-task prospective), the lighter colour represents results from the present experiment (single-task prospective).

To investigate whether the duration ratings were higher in a single-task prospective paradigm than in a retrospective paradigm, the duration ratings over items were compared between the present study and the retrospective study reported in Experiment 2. Results from a repeated measures ANOVA with duration ratings aggregated up to the item level as the dependent variable revealed that that

there are was a significant main effect of experiment ($F(1,81)=14.15$, $p<.001$, $\eta_p^2=.15$). This suggested that ratings in the single-task prospective paradigm ($M=4.21$, $SE=.09$) were significantly higher than ratings in the retrospective paradigm ($M=3.99$, $SE=.10$).

4.9.4 Regression results

To further investigate the independent contribution of the number of perceived event boundaries and the similarity between them on the duration judgements, by-item hierarchical multiple regression analyses were conducted to examine the proportion of variance accounted for by the similarity between events and the number of segments as measured by the web questionnaires (see section 2.7.3) over and above the actual clock duration of the animations. Clock duration was added to the model first as a control predictor, as it accounts for the systematic variation between the triads. The results showed that when the number of segments was added to this regression model, the proportion of variance accounted for increased significantly from .52 to .62 ($F_{\text{change}}(1, 81)= 14.92$, $p<.001$). When similarity between events was added to the latter model, the proportion of variance accounted for did not increase significantly (remaining at .62, $F_{\text{change}}(1, 80)= .004$, $p=.95$) (Table 7). Changing the order in which clock duration, the number of segments and similarity between events were added to the model did not affect the pattern of significance. This suggests that there is a significant role for the number of sub-events, but not for similarity between sub-events in explaining the duration estimates.

Table 7. Regression coefficients for duration ratings. Note: ** indicates $p \leq .001$, * indicates $p < .05$.

	Model	<i>B</i>	<i>SE B</i>	β
1.	Constant	2.93	.25	
	Clock duration	.21	.04	.52**
2.	Constant	2.73	.24	
	Clock duration	.09	.05	.21*
	Number of sub-events	.21	.05	.45**
3.	Constant	2.57	.39	
	Clock duration	.09	.05	.21*
	Number of sub-events	.21	.05	.45**
	Similarity	-.004	.06	-.005

4.10 Discussion

The results presented here suggest that as expected, there was no significant effect of condition on the response latencies for the recognition memory task in a single-task prospective paradigm. Furthermore, there appeared to be no significant effect of condition on the duration ratings obtained in the duration rating task. However, the results from the regression analysis suggest that even though there is no significant effect of condition, the number of perceived segments (but not perceived similarity) explained a significant proportion of variance over and above clock duration. These findings suggest that event boundaries may serve as

temporal markers, as the number of segments positively affects duration judgements, even when the content of the events themselves does not need to be remembered.

Furthermore, comparisons between experimental paradigms suggest that there was no difference between the single-task prospective paradigm and the dual-task paradigm above in terms of response latencies and duration ratings (this will be further discussed below). Comparisons with the retrospective paradigm in Experiment 2 suggest that recognition times are faster and duration ratings are higher in a prospective paradigm.

4.10.1 Event boundaries as temporal markers

These results are partly in line with findings from previous research showing an effect of the number of events on prospective duration estimates. The current data show an effect of number of segments in the regression analyses, but not between conditions. This suggests that over all, the number of segments might be employed to inform the time monitoring process, but that perhaps more subtle manipulations of the number of sub-events such as those between conditions may not affect the boundaries that are used. The boundaries between events that are relevant for remembering the amount of time may thus be employed in the time keeping process, suggesting that perhaps not all boundaries are equally relevant for this task.

Given that event segmentation is a hierarchical process, the lack of an effect of condition gives rise to the idea that coarse rather than fine-grained segments may guide time keeping. On a very coarse level, all animations within a triad are similar in terms of event structure: the animations are the same in terms of movement (speed, trajectory, etc.) of the main protagonist(s), providing each animation with the same ‘rhythm’ or temporal coherence (cf. Boltz, 1995). Moreover, as shown by Zacks’ research, coarse boundaries correlate with some fine-grained units in such a way that coarse segments are made up of groups of fine-grained segments (Zacks, 2004). Thus, the relationship between number of

event boundaries and prospective duration estimation may be due to the way in which the coarse event structure imposes a temporal rhythm on the stimulus, facilitating time keeping. The difference in number of sub-events at a fine-grained level may be more salient in a memorising task, where the identity of the segments matters, than in a time keeping task, leading to an effect of quantity in the regression analysis, but not an overall difference between conditions.

This could also explain the lack of an effect of similarity on the duration estimates: if only the very coarse event structure is employed, differences in similarity may not be relevant to the time keeping process, as similarity as manipulated in this study does not directly contribute to time keeping. Given that participants are not asked to remember the content of the stimuli, fine-grained information about (sub-)event boundaries may not be relevant to the time keeping task and thus not remembered. Thus, an effect of similarity may rely on remembering content. This will be further discussed in section 4.11.

4.10.2 Recognition memory and response latencies

Contrary to the previous experiments, the response latencies in the recognition memory task do not appear to show a linear trend over conditions. This could be due to the fact that the over all, the same coarse information was associated with all conditions of a triad, so there is no difference in information associated with the probes across conditions. A potential confound however is that the criteria for excluding participants based on the memory task were less stringent (overall recognition accuracy $\leq 50\%$ rather than a more stringent recognition accuracy $\leq 50\%$ in one of the conditions). Overall recognition accuracy was lower for this experiment than for the other experiments, which was expected, as participants did not explicitly study the content of the animations. As the memory task may have been more difficult for this group of participants, the previous exclusion criteria seemed too stringent, as these would lead to the exclusion of a large number of participants (11 out of 79). However, reanalyses of the data with these stringent exclusion criteria (excluding 11 participants based on these criteria, plus

5 more participants for the purpose of counterbalancing, excluding the participants with the next worst memory performance for those lists, leaving a sample of 63 participants in total) show that results remain identical, with no significant effect of condition for the response latencies or the duration ratings, and the same pattern of results for the regression analyses. Although statistical power is lower for this smaller sample, these results suggest that there appears to be no effect of condition, even when participants displayed good memory performance ($M=85\%$, $SD=9.3\%$; basic: 84%; numerous: 87%; dissimilar: 83%).

Furthermore, response latencies in the single-task paradigm did not differ from those in the dual-task paradigm. This does not concur with the prediction that response latencies should be slower when less information (i.e. only duration and not content) is encoded about the stimuli. However, the accuracy scores in the present paradigm were lower than those for the dual-task paradigm. This suggests that there could have been a speed-accuracy trade-off: participants could have been faster because they were less accurate. Another option is that because content was not encoded, these recognition decisions were based on familiarity only (not involving any recollection of content), suggesting that the recognition memory task in Experiment 6 might have relied on a different process altogether than the recognition memory tasks in the previous chapters (i.e., familiarity-based recognition versus recollection; see Yonelinas, 2002 for an extensive review) and might thus not be comparable. Further research could investigate the relationship between encoding different types of information about stimuli and recognition speed and accuracy in more detail.

4.10.3 Working memory versus episodic memory

The current results suggest that event boundaries may serve as temporal markers. This is in concurrence with the findings from Poynter and Homa (1983) and Waldum and Sahakyan (2013), but the pattern of results in the present study is in the opposite direction to that of Liverence and Scholl (2012): more events predict longer duration ratings over and above actual duration. As pointed out above, this

discrepancy could be due to the difference in task and measures used, hence, a difference in cognitive mechanisms that are involved in the task.

In their task, Liverence and Scholl asked participants to provide a duration reproduction immediately after viewing the stimulus. Thus, participants kept the stimulus duration in their working memory. Working memory may be limited in the amount of information that can be held and hence, information can be flushed from working memory. Liverence and Scholl argued that the flushing of an event based memory buffer (Kurby & Zacks, 2008) explains why the duration ratings for their trials with more events are reproduced as shorter than their counterparts with fewer events. The present results do not dispute this notion, but suggest that these findings do not generalise to other estimation methods that easily: if we witness an event and know that we will be asked to estimate its duration at a later point in time, it is likely that we do not keep the duration of the events in working memory. Rather, we store a representation of the events in episodic memory (that may or may not include information from time keeping processes) that together with our knowledge of events in the world could form the basis for a duration estimate.

Furthermore, one can ask whether a working memory model that relies on buffering holds for longer durations. Liverence and Scholl claim that the same mechanisms are at play when we perceive time in naturalistic events, for example a day at an amusement park. Time appears to go quickly, which according to Liverence and Scholl is due to the high number of events and the coarse segmentation. However, we argue that this is a measure of passage of time, rather than duration. Estimating the duration of a day at an amusement park is unlikely to rely on a working memory representation, but rather, it is likely to be based on an episodic representation of the events that took place. We therefore argue that working memory may play a role in remembering short durations for immediate recall, where stimulus presentation rate is stored, and that the present results suggest that this may only be the case when participants need to store the information in (visual) working memory for immediate recall. The present results suggest that an increase in the number of events is expected to lead to an increase

in duration rather than a decrease for tasks where the duration is stored in episodic memory, as suggested by the positive relationship between duration ratings and number of perceived segments.

4.10.4 The effect of a secondary task

Contrary to our hypothesis that duration ratings in a single-task would be higher due to more attention to time, the comparison across Experiment 5 and 6 revealed no significant difference between the duration ratings between both experiments when taking all conditions into account. However, there appears to be evidence for a trend where the basic condition is rated as longer in the single-task prospective paradigm than in the dual-task. Given the nature of the secondary task, a difference between the experiments for the other conditions could be obscured: in the dual-task paradigm, ratings for the numerous and dissimilar conditions are higher because of the increase in remembered event structure. Hence, remembering content as a secondary task is likely to obscure general differences between a single-task paradigm and a dual-task paradigm, as remembering content gives rise to an increase in duration ratings for the numerous and dissimilar conditions. Further research could investigate this using a secondary task that does not require participants to focus on the content of the animations, shedding light on whether duration estimates indeed are shorter when there is a secondary task. This would be in line with findings from previous studies discussed above (e.g. Block & Zakay, 1997).

4.11 General discussion

4.11.1 The role of memory in prospective paradigms

Experiment 5 and 6 provide evidence against the prediction from Block, Hancock and Zakay (2010), who suggested that prospective judgements would not be affected much by encoding information for later tasks. The studies above have shown that whether or not participants remember the content of the stimuli for a

subsequent task greatly affects the pattern of duration estimates. When participants encode stimulus content for a later task, there is an effect of condition, and the number of perceived event boundaries and the similarity between them predict the duration estimates. However, when participants do not encode this information, there is only an effect of the number of (coarse) event boundaries on duration ratings. Thus, the current experiments present evidence against this claim by showing that deeper encoding of content does affect prospective duration estimation.

4.11.2 The role of event structure in prospective paradigms

Both experiments in this chapter have suggested that event structure plays a role in prospective paradigms. When participants encode the stimuli for later recall (Experiment 5), the pattern of results is similar to that in a retrospective paradigm. As illustrated by Experiment 6, the effect of the number of event boundaries as shown by the regression analysis might be due to the effect of coarse event structure on time keeping. Additional studies manipulating the event structure on a very coarse level could further investigate whether indeed changes in the number of coarse boundaries affect prospective duration estimates, and under what task conditions. There is some evidence available from previous studies that indeed the ease of encoding of the temporal unfolding of events affects duration estimates: continuous events (e.g. throwing a ball) and coherent events (e.g. dribbling with a basketball) elicit more accurate duration estimates than incoherent events (e.g. dribbling with uneven periodicities), showing greater overestimation for incoherent events in both prospective and retrospective paradigms (Boltz, 2005). Thus, investigating what boundaries are relevant for a time keeping task (i.e. how coarse or fine grained) and how their quantity and periodicity affects duration estimation appears to be an interesting avenue for further research.

Together, the findings from the studies presented here suggest that whether or not a modulatory effect of event structure can be observed in prospective paradigms depends on the type of event boundaries that are manipulated, and on

whether or not participants are instructed to remember the content of the stimuli. This also leads to predictions about whether or not event complexity plays a role in prospective paradigms. Given that the current data only showed an effect of similarity when participants actively remembered the content of the stimuli, a prediction for further research could be that stimulus dissimilarity only plays a role when content needs to be remembered. A second prediction would be that the content of stimulus segments (e.g., complexity or similarity) could play a role even when content does not need to be remembered, but only when the complexity somehow imposes a different coarse temporal segmentation by providing different event boundaries than a less complex stimulus. This would however make it a quantity effect rather than a quality effect in terms of qualitative differences between sub-events. The fact that there has been little agreement in the literature on whether there is an effect of complexity in prospective paradigms could (at least partly) be explained by these observations on task- and stimulus-relatedness of the effect and therefore these predictions could form a useful basis for further research into the effect of complexity.

4.11.3 The effect of a secondary task in prospective paradigms

Previous studies had already shown an effect of secondary task on duration estimates in prospective paradigms, with more difficult secondary tasks leading to shorter duration estimates. The current studies are in partial agreement with these results, because as shown by Experiment 5, the nature of the secondary task (i.e., remembering content) matters. Experiment 5 showed that when participants remember stimulus content as a secondary task, duration ratings do not become shorter over all, but do show an effect of condition, precluding any detrimental effects by showing an increase for more remembered stimulus information. Dual task paradigms are detrimental to performance when different stimuli are processed in parallel (e.g., words and songs) or different aspects of the same stimuli are processed (monitoring words starting with a letter in reading a story), which required divided attention. This chapter has argued, in contrast, that certain

stimulus characteristics contribute to the prospective duration estimation process. The approach taken here differs from previous studies in that the same stimulus is used for the time keeping task and the content memory task. These two tasks are related to each other, in fact leading to more encoded information about the animation (duration and content). The current study contributes to the literature by showing that attending to related aspects of the same stimuli elicits an effect of event structure.

4.12 Conclusion

In sum, the results presented here suggest that event structure plays a role in prospective duration estimation. When participants pay attention to both time and content, duration judgements are modulated by the number of identified event boundaries and the similarity between them. When participants only pay attention to time, duration judgements are only modulated by the number of (coarse) event boundaries. Together these experiments provide support to the idea that both the number of segments and the underlying similarity structure affect duration judgements, but that similarity only plays a role when participants have a deeper encoding of the event structure, whereas the number of segments also plays a role when participants only pay attention to duration.

Chapter 5

The effect of event structure on reconstructed duration: a neuroimaging study

5.1 Introduction

As argued in the previous chapters, time and memory are two interconnected concepts. Tulving's theory of episodic memory (Tulving, 1972, 1984, 2002) states that episodic memory critically relies on the temporal organisation of memory, in which events can temporally precede, succeed or overlap. Time, therefore, can be seen as the organisational principle of episodic memory. Tulving (1985) observed that amnesic patient K.C. was not able to mentally travel through time, for instance go back to his own past or think about his future. Dissociations with other processes of memory such as semantic memory (or knowledge) suggest that the ability to remember the past and reason about the future is thus a key characteristic of episodic memory (Tulving, Schacter, McLachlan, & Moscovitch, 1988), and that impairments in episodic memory are associated with lesions or atrophy in the hippocampus and adjacent cortex (Rosenbaum et al., 2000). A growing body of research in both animals and humans suggests that indeed the hippocampus and surrounding areas in the medial temporal lobe are critical to representing the temporal organisation of memories (Eichenbaum, 2014; MacDonald, 2014). The following sections give a brief overview of studies that have investigated the role of hippocampus in episodic memory, focusing on its role in recollection, sequence memory and temporal representation. Furthermore, this chapter will argue that hippocampus is a likely candidate for being involved in retrospective duration estimation, which is corroborated in the fMRI study presented here.

5.2 Hippocampus in episodic memory and time

5.2.1 Medial temporal lobe and its role in episodic memory

The human medial temporal lobe (MTL) is a set of areas of the brain that comprises the hippocampus, its adjacent cortex (the parahippocampal cortex, the entorhinal cortex, and the perirhinal cortex), the fusiform gyrus and in some frameworks, the amygdala (Squire & Zola-Morgan, 1991; Amaral, 1999). It has long been argued that the MTL plays a role in encoding events into memory. In the 1950s, Scoville and Milner (1957) observed a selective loss of recent memories after removal of both MTLs in patient H.M. due to severe epilepsy. This insight has led research to focus on the role of each of the sub-regions of the MTL in memory formation. In particular, many studies since have attempted to unravel whether recollection of episodic memories and familiarity-based recognition are supported by distinct neural underpinnings, or whether both rely on a single brain process, and what the role of different sub-regions of MTL is in both kinds of memory (Eichenbaum, Yonelinas, & Ranganath, 2007).

5.2.1.1 Medial temporal lobe and its role in recollection

Lesion studies in amnesic patients have indicated that in particular the hippocampus is selectively involved in representing and recollecting associations between items in memory and contextual associations, rather than in recognising single items based on familiarity (summarised in: Eichenbaum et al., 2007). For example, Yonelinas and colleagues (2002) investigated recollection and familiarity in patients who suffered from mild hypoxia following cardiac arrest. Hypoxia is the temporary loss of oxygen, which can result in hippocampal atrophy while leaving surrounding cortex intact. These patients showed severely impaired recollection, but normal familiarity-based recognition. Control participants with more extensive damage to the MTL showed deficits in both recollection and familiarity-based recognition, suggesting that although MTL plays a role in both recollection and

familiarity-based recognition, hippocampus appears to be particularly involved in recollection.

Furthermore, studies using fMRI in healthy individuals have shown that the hippocampus plays a selective role in the retrieval of episodic but not non-episodic memories. For instance Eldridge and colleagues (2000) asked participants to memorise each presented word for a subsequent memory task. In the fMRI scanner, they were exposed to new and studied words and were asked whether they remembered studying the word (i.e. if they could recollect the moment that they studied the word during the previous task) or whether they just 'knew' the word if they could not recollect this (a task based on Yonelinas & Jacoby, 1995). The results show that the hippocampus is selectively involved in retrieving episodic memories (i.e. recollection), but not non-episodic memories (i.e. familiarity-based recognition). Thus, both lesion studies and fMRI studies suggest that the hippocampus is involved in recollecting episodic memories, rather than making familiarity-based judgements.

Furthermore, patients with damage in the MTL typically show impaired relational memory. Even when item recognition (i.e. familiarity) is matched, patients with MTL damage are worse at identifying when or where an item was presented, which modality it was presented in, which context it was presented in or how frequently it was presented (reviewed in: Yonelinas, 2002). In particular, lesion studies have shown that damage to the hippocampus results in deficits in memory for temporal order (e.g. which list was an item presented in, what order were items presented in) (Mayes et al., 2001) and associations between objects and location (Vargha-Khadem et al., 1997). Using fMRI, studies in healthy participants have shown that activity in the hippocampus and surrounding parahippocampal areas is associated with the strength of encoding of episodic memories: for instance, activity in these areas during encoding is correlated with the accuracy of later associative memory for context or source during recollection (Davachi, Mitchell, & Wagner, 2003). Together, these findings suggest that hippocampus plays a role in remembering contextual information about events, including information about temporal development.

5.2.1.2 The role of hippocampus in thinking about events

The evidence summarised above suggests that the MTL and particularly hippocampus plays a role in remembering and recollecting associative information about stimuli, such as associations between items, and their context in terms of space and time. As outlined in Chapter 1, event representations are thought to consist of these binding relationships between entities and their temporal and spatial location. The hippocampal memory system has been implicated in binding together these separate elements that make a memory, and is thought to provide a 'conjunctive code' that describes how we segment and relate information in memory (summarised in: Shastri, 2002). When an element of an event is probed, the hippocampal system is thought to complete the event by retrieving the other attributes that are encoded together through this conjunctive representation (Shastri, 2002).

There is a large body of research investigating the role of the MTL in remembering past events and imagining future events. The first study to investigate the role of hippocampus in thinking about past and future events used positron emission tomography (PET) to investigate whether retrospective memory and prospection rely on the same neural networks, showing that both processes involve medial temporal areas (Okuda et al., 2003). More recently, neuroimaging studies using fMRI have suggested that indeed hippocampus plays a role in re-experiencing past events and pre-experiencing future events. For example, Botzung and colleagues have shown that left hippocampus activation can be observed during thinking about the past and during thinking about the future, suggesting that this area plays a role in (re-)constructing event representations (Botzung, Denkova, & Manning, 2008). Addis and colleagues also observed activity in the left hippocampus in both thinking about past and future events, and additionally observed that particularly left hippocampus appears to be involved in past event reconstruction, and that both hippocampi play a role in mental elaboration of events (Addis, Wong, & Schacter, 2007). In a further analysis of the same data, Addis and Schacter (2008) found that the amount of detail that is retrieved about

past and future events modulates activity in left posterior hippocampus: more recalled detail leads to more activity. These findings provide evidence for a role for hippocampus in retrieving episodic details about (past) events that can be used for future simulations (Addis & Schacter, 2008).

Further studies investigating the role of the MTL in constructing and reconstructing events have shown that hippocampus plays a key role in (re)constructing scenes and events, both when these are real personal experiences and when these are new fictitious experiences (Hassabis, Kumaran, & Maguire, 2007). Furthermore, it was found that in patients with hippocampal damage, the ability to imagine new fictitious experiences is impaired, suggesting that indeed hippocampus plays a critical role in this process (Hassabis, Kumaran, Vann, & Maguire, 2007). Together with many other studies, these studies suggest that hippocampal areas support the construction of future events by recombining (event) information into a representation or simulation of new events (Schacter, Addis, & Buckner, 2007; Schacter & Addis, 2009). Constructing an event representation requires the retrieval of conceptual and visuospatial information, and this information needs to be bound into a coherent event. Given the involvement of hippocampus in building associative memories by binding relational and contextual information, it is thus not entirely surprising that the hippocampus plays a role in binding event information for novel or future events (Addis et al., 2007). Hippocampal engagement has also been shown for events that are novel but 'atemporal' in that they are not located at a specific point in the past or future, suggesting that hippocampus is indeed involved in binding event information regardless of temporal direction (Hassabis, Kumaran, & Maguire, 2007).

Hippocampus has also been implicated in binding event information *across* space and time. Using a paradigm in which representational gaps needed to be bridged across space (e.g. an object and its colour presented in a spatially discontinuous way) or time (e.g. object and colour presented in a spatiotemporally discontinuous way), Staresina and Davachi (2009) have shown that activity in hippocampus increases as a function of spatiotemporal discontinuity: when

comparing against non-discontiguous stimulus presentation (e.g. an object of a specific colour), hippocampal activity increases significantly when object and colour need to be bound across space (i.e. different locations in screen). Over and above that effect, hippocampal activity again increases significantly when object and colour need to be bound across space and time, suggesting that activity in the hippocampus is modulated by increasing demands to integrate disparate associative information. These findings emphasise the unique ability of hippocampus to flexibly integrate contextual and associative information into an event representation.

Furthermore, the MTL appears to be more active when object information that is associated with an event boundary is retrieved. In a study by Swallow and colleagues (2011) participants watched videos and were then faced with a two-alternative forced choice recognition task, asking them which of two objects was in the movie they just saw. Their results indicated that hippocampal and parahippocampal areas were more active when object information needed to be retrieved across an event boundary than within an event boundary. According to Event Segmentation Theory, an increase in perceptual processing is to be expected when an event boundary is perceived, and the mental model of the situation needs to be updated in order to represent the current situation. In that case, objects that are associated with event boundaries are more likely to be encoded into episodic memory than objects that are not associated with a boundary (Swallow et al., 2011). In addition, the retrieval of object information across an event boundary depends on episodic memory rather than working memory, as the working memory buffer is cleared when an event boundary is encountered. The results indeed indicated that the episodic memory network was more involved in retrieving information across an event boundary than within an event, suggesting that information across an event boundary needs to be retrieved from episodic memory. This may also lead to stronger memory representations, as the relevant object information has been stored in episodic memory and may be less susceptible to memory interference (Swallow et al., 2011). Together, this evidence suggests that an increase in hippocampal activity is to be expected when

information associated with event boundaries needs to be retrieved, and when information needs to be retrieved across event boundaries.

In order to reconstruct an event from memory, one needs to have a representation of its content and information about the relevant event boundaries, and these representations need to be anchored in time and space. As argued above, representations of events consist of multimodal information about the content of events: for instance entities and their properties (e.g. colour, size) and information about their motion and trajectory can be part of an event representation. Thus hippocampus is thought to bind information supplied by corresponding sensory-motor regions, such as the visual information processed by a subsystem of visual areas. Previous studies have found a network of posterior visual regions co-activated with hippocampus that is involved in remembering past events that includes the lingual, occipital and fusiform gyrus (Slotnick & Schacter, 2006; Addis et al., 2007). These regions are thought to be part of a functional network that includes these visual areas as well as the posterior portion of the parahippocampal cortex and the hippocampus (Addis et al., 2007; Kahn, Andrews-Hanna, Vincent, Snyder, & Buckner, 2008). This network has been shown to form a functionally distinct cortical pathway that converges in the hippocampus, and has been implicated in tasks such as recollection, autobiographical memory and event (re)construction (Kahn et al., 2008). During retrieval of events, this network of domain-specific cortical regions is often reactivated together with activity in hippocampus (Buckner & Wheeler, 2001; Slotnick & Schacter, 2006; Addis et al., 2007).

5.2.1.3 The role of hippocampus in temporal coding

Episodic memory relies on the binding of stimuli to their spatial and temporal context, connecting what happened where and when (Eichenbaum, 2004). For long, hippocampus was thought to mainly contribute to constructing episodic memories by providing a cognitive map that defines a spatial context. It was observed that particular locations in the environment are encoded by 'place cells'

that selectively fire when an animal is in a specific spatial context (for more information on place cells, see O'Keefe & Nadel, 1978 and Burgess, Maguire, & O'Keefe, 2002). As discussed above, however, evidence suggests that the hippocampus is also involved in encoding temporal context. As outlined below, neurophysiological studies have recently discovered hippocampal cells that (also) fire in response to temporal context that, in analogy with 'place cells', have been deemed 'time cells'. As discussed below, previous studies have identified two types of temporal information – temporal anchors and temporal order – that are represented by hippocampal neurons. Note that the evidence mainly comes from animal studies – to date, only a handful of studies have reported on such data from humans. The following will give a brief overview of the role of hippocampus in representing temporal context in both animals and humans, focusing on evidence for the existence of temporal representations in hippocampus, and how such information might be integrated by the hippocampus into a temporal context.

The earliest evidence for the existence of time cells comes from studies that investigated freely moving rats performing a task in which they were required to remember the temporal order of a sequence of odour stimuli, receiving a reward for identifying which odour was presented earlier. The results showed that stimuli that occurred closer together in time were represented more similarly to one another by hippocampal ensembles than stimuli with greater temporal distance to one another, suggesting that temporal context is represented by 'time' cells in the hippocampus (Manns, Howard, & Eichenbaum, 2007).

Similar evidence was obtained by MacDonald and colleagues who investigated the relative contribution of different variables including location, speed, head direction and time to hippocampal firing patterns (MacDonald, Lepage, Eden, & Eichenbaum, 2011). Observing these patterns in freely moving rats performing a delayed association memory task (pairing objects and odours), they found that passage of time is encoded in the hippocampal firing pattern. Furthermore, they found that place and time together best predicted the pattern of activity in hippocampal cells, suggesting that indeed hippocampus plays a role in representing spatio-temporal context. In a follow-up study, they found that some

cells fired relative to a time point in the past, suggesting a potential role in encoding retrospective duration. Likewise, some cells showed an increase in firing rate towards the end of an interval, potentially anticipating an endpoint or upcoming decision.

Furthermore, MacDonald and colleagues observed that most hippocampal cells started firing in response to a new temporal correlate when the temporal context changed. Importantly, based on these findings, MacDonald and colleagues proposed a mechanism by which the hippocampus encodes the when and where of events. Hippocampal cells perform partial remapping and retiming when there are changes in location and time: most of the spatial and temporal correlates of hippocampal cells are updated, but some firing patterns remain anchored to their old correlate. This pattern of partial remapping was proposed to be fundamental to encoding relational information between current and previous location and point in time, and suggests that the hippocampus plays a role in representing events by binding event sequences across space and time.

Similar results were found in primates. Investigating the contribution of individual structures in the MTL to temporal order memory, Naya and Suzuki (2011) recorded the neurophysiological activity (spiking) of single cells in hippocampus, entorhinal and perirhinal cortex, and from a visual area (TE) in monkeys that performed a temporal-order memory task in which they were required to encode two visual items and their temporal order. Results showed that activity in hippocampal time cells during encoding is incremental in relation to the relative time between the current moment and the last cue presentation, and between the current moment and the next cue presentation. The authors suggested that this timing signal might serve to anchor the timing of events in episodic memory. Furthermore, they proposed that this information is propagated via entorhinal cortex to perirhinal areas, where it is integrated with item or object information from visual areas to obtain a representation of the sequential temporal order of items or events.

More evidence for the role of time cells in remembering temporal anchoring (i.e. when and how long ago an event occurred) comes from a study by Mankin and

colleagues (2012) in freely moving animals. Rats were exposed to the same event (foraging for food) at different intervals. Results suggested that hippocampal representations varied more between events that are further apart in time (6 hour interval) than events that are closer (1 hour interval), suggesting that temporal distance is encoded by hippocampal neurons. Together, these findings suggest that hippocampal cell ensembles encode both the temporal order of different events, as well as the temporal distance between events.

Furthermore, there is evidence that the hippocampus and surrounding areas play a role in temporal representations humans. For example, Kumaran and Maguire (2006) used fMRI to explore the role of hippocampus in detecting anomalies in temporal order. They found that when the order of a sequence of presented images has been changed when the same images are presented again, activity in left hippocampus and surrounding entorhinal and perirhinal regions increased, suggesting that hippocampus is involved in generating predictions about the unfolding of events, and in monitoring violations of these predictions. Furthermore, this increase in activity was only observed when a new order of already presented events was presented, and not when participants encoded a sequence of new events, suggesting that indeed hippocampus plays a role in maintaining a representation of the temporal order of events.

More direct evidence for the role of hippocampus in encoding temporal order in humans comes from a study by Paz and colleagues that investigated single cells in patients with pharmacologically intractable epilepsy who as part of their treatment were implanted with chronic depth electrodes to determine the locus of their seizures (Paz et al., 2010). In these patients, it is thus possible to investigate the response of single cells and cell assemblies, which is not possible with other neuroimaging methods such as fMRI or PET due to limitations in spatial and temporal resolution. Therefore, this study has been crucial in translating findings from animal research to human neuroscience. In the experiment, subjects were asked to watch a series of short cinematic clips (e.g. famous people or characters, animals in motion, famous landmarks or sceneries, objects in dynamic context) several times without any additional task. The results from single cell recordings

indicated that the pattern of neuronal activity during successive segments gradually became more correlated, suggesting that hippocampal neurons create or encode temporal associations across events, and that a temporal representation emerges over subsequent stimulus exposures. These findings demonstrate that also in humans, there appears to be a neuronal correlate in the hippocampus for building representations of temporal relationships between successive events.

Recently, several studies have investigated the nature of these temporal representations using novel fMRI analysis methods such as multi-voxel pattern analysis. These studies have shown that multivariate patterns of hippocampal activity contain information about the temporal context of objects. For example, Schapiro and colleagues (2012) have shown that the representational similarity between objects that are nearby in time increases: two presented visual objects (in this case, fractals) will exhibit greater representational similarity when they are presented sequentially with high regularity than objects that are presented with lower sequential regularity. Further evidence for the role of the MTL in extracting such temporal regularities comes from a clinical case study showing that bilateral hippocampal loss and broader MTL damage leads to impaired extraction of temporal regularities (i.e. impaired statistical learning) (Schapiro, Gregory, & Landau, 2014). In a familiarity judgement task, the patient performed at chance, showing no differentiation between sequences of items with high co-occurrence during the exposure phase and items that had no co-occurrence. These findings suggest that indeed hippocampus and surrounding areas in the MTL are critically involved in remembering sequences and extracting information about regularities in sequences.

Further investigating encoding and retrieval of object sequences, a very recent study by Hsieh and colleagues (2014) has indicated that the temporal position of objects in a learned sequence is encoded in hippocampal voxel patterns. In left hippocampus, these patterns appear to be sensitive to sequence boundaries, as pattern similarity was greater for within-sequence than between-sequence pairs (Hsieh et al., 2014). These findings have been corroborated by a very similar (concurrent) study by Ezzyat and Davachi (2014), suggesting that indeed the

representational similarity between objects in left hippocampus may be indicative of their temporal relationship, and that this mechanism may be crucial for the organisation of temporal memory.

Note that hippocampus has not been implicated in timing tasks per se. A meta-analysis of human fMRI studies using supra-second perceptual timing tasks does not report activity in any of the MTL areas (Wiener, Turkeltaub, & Coslett, 2010). Instead, the meta-analysis has shown that areas such as the supplementary motor areas and inferior frontal gyrus appear to be implicated in timing tasks. Furthermore, sub-second timing tasks have been shown to elicit additional activity in basal ganglia and thalamus, which is in concordance with internal clock models of very short interval timing that propose a relationship between sub-second timing and striatal beat frequency (Matell & Meck, 2004). So, although there appears to be a different network of regions involved in active time keeping, hippocampus and surrounding MTL regions are thought to play a role in representing temporal anchoring of events and sequential representation. Rather than being an internal clock, hippocampus appears to play a role in representing temporal associations across events and binding event information together into an event representation.

5.2.1.4 The role of hippocampus in thinking about temporal aspects of past events

Very few studies have investigated the relationship between activity in the hippocampus during encoding and the retrieval of temporal memory in humans. Following up on the question of what the role of MTL is in encoding and retrieving temporal information, Tubridy and Davachi (2011) have shown that the magnitude of activation in the hippocampus during the encoding of a word triplet correlates with later memory for its temporal order. This suggests that greater activation of hippocampal and parahippocampal regions during encoding leads to a more accurate memory for temporal order. These findings indicate that hippocampus and parahippocampal regions may play a role in encoding temporal sequence

information that underlies episodic memory formation, which enables successful retrieval of this information.

However, this Turbidity and Davachi did not explicitly investigate whether hippocampal and parahippocampal regions are involved in the retrieval of temporal information, as opposed to encoding. In one of the few studies investigating the role of MTL in retrieving temporal information, participants were asked to drive around and pick up passengers in a virtual reality environment, followed by an fMRI scan during which they were asked to indicate which passenger they had encountered first (Ekstrom & Bookheimer, 2007). The results indicate that compared with brain activity elicited by a spatial recognition task (e.g., “have you seen this before”), activity in hippocampus was larger for temporal order retrieval. Vice versa, activity in parahippocampal areas was larger for spatial retrieval.

Another study investigating the role of the MTL in retrieving temporal information was conducted by Lehn and colleagues (2009), who observed both hippocampal and parahippocampal contributions to the retrieval of temporal order. In their experiment, participants were asked to watch a novel movie and remember as much of it as possible, followed by an fMRI scan the consecutive day during which they were asked to reconstruct the temporal order of the movie by rearranging and mentally replaying scenes from the movie in the correct order. A region of interest analysis of MTL regions showed that an increase in activity in hippocampus and parahippocampal areas can be seen during the retrieval of temporal order information. This effect in the MTL was only observed for reconstructing the temporal order of remembered events; there was no involvement of MTL in the construction of temporal order information of novel sequences based on logical rules (sequential reasoning). These findings suggest that activity in hippocampus and parahippocampal areas is specific to the retrieval of temporal information about previously encoded events. In sum, the few studies that have looked into the role of hippocampus in retrieving temporal aspects of events suggest that hippocampus indeed appears to be involved in retrieving information about the temporal unfolding of remembered events.

5.3 Experiment 7: The role of hippocampus in retrospective duration reconstruction.

5.3.1 Research question and aims

Taking into consideration the evidence discussed above, hippocampus is likely to play a role in providing a temporal framework for the organisation of episodic memory. More specifically, left hippocampus may play a critical role in representing the temporal development of events. Firstly, hippocampal representations, and in particular those in left hippocampus, contain information about the temporal order of (sub-)events (cf. Kumaran & Maguire, 2006; Hsieh et al., 2014; Ezzyat & Davachi, 2014). In order to remember or reconstruct temporal development, one needs a representation of the sequential relationship between sub-events. This is in line with theories about the nature of episodic memory: episodic representations rely on having a representation of whether events preceded or succeeded each other, or whether they co-occurred in time.

Secondly, hippocampus appears to also be involved in representing information about temporal anchoring. Temporal anchoring may be critical to retrospective duration estimation, as it provides information about the beginning and end of perceived events. Note however that temporal anchors may allow us to retrospectively *estimate* duration, rather than providing an explicit, absolute measure of temporal duration (in seconds, minutes or hours). Hippocampal representations thus contain information about the ordinal relationship between events as well as information about how much time has passed between them (MacDonald, 2014).

As outlined above, very few studies have investigated the role of hippocampus in thinking about temporal aspects of past events. In particular, none of these studies have addressed whether the structure of the perceived events is reflected by hippocampal activity during duration estimation. The present study therefore examined whether the content of perceived events, and therefore the content of episodic memory, in terms of number of perceived events and similarity

between them affects hippocampal activity during retrospective duration estimation. Based on the role of hippocampus in representing information about the order of events (in particular, left hippocampus, cf. Kumaran & Maguire, 2006; Hsieh et al., 2014; Ezzyat & Davachi, 2014) and their temporal anchoring, duration estimation-related activity in (left) hippocampus would be expected to be affected by the number of perceived sub-events. Furthermore, activity in the hippocampus has been shown to increase when information about objects associated with event boundaries is retrieved and when information has to be retrieved across event boundaries (cf. Swallow et al., 2011). Thus, we expected to find an increase in activity in hippocampus when information about more event boundaries needs to be retrieved.

Furthermore, following the reasoning outlined in previous chapters, activity in hippocampus may also be affected by the similarity between sub-events. Previous studies have indeed found that left hippocampus is involved in reconstructing representations of past events (Addis et al., 2007; Botzung et al., 2008), and that activity in left posterior hippocampus increases as an effect of the amount of episodic detail that is retrieved (Addis & Schacter, 2008). Given that retrospective duration estimation appears to rely on reconstructing duration based on event structure and event content, events with lower similarity between sub-events require more episodic detail to be encoded in order to obtain an accurate representation. Therefore, we expected to find an increase in activity in left hippocampus as an effect of the amount of event information that needs to be retrieved to inform the duration estimation process, with increasing activity for more dissimilar events.

Although (left) hippocampus appears to be a likely candidate for involvement in retrospective duration estimation, to the author's knowledge, studies up until now have not investigated retrospective duration estimation using neuroimaging methods. Given the limitations of a retrospective design (i.e., naïve participants, limited number of items; see Chapter 2), previous studies have not been able to or have not attempted to investigate retrospective duration estimation using fMRI, probably due to limitations regarding the statistical power of the

design. For example, many behavioural studies have used one-off retrospective judgements (i.e. one judgement of one interval of time), which would not provide sufficient fMRI time-series data to perform statistical analyses on, as many repetitions of stimuli are necessary to reliably detect the stimulus induced BOLD-signal (i.e. low signal-to-noise ratio; Huettel, Song, & McCarthy, 2009).

The present study aimed to address this hiatus by using the paradigm that was developed and employed in the previous chapters (in particular in Experiment 2, Chapter 2) in the fMRI scanner. This novel retrospective duration estimation uses 30 items (28 stimuli plus 2 anchors), allowing us to obtain 28 responses per participant (or 8-9 responses per condition per participant). In the present study, this paradigm was used in the fMRI scanner to investigate the neural correlates of retrospective duration estimation, particularly focusing on whether hippocampus is indeed involved in retrospective duration estimation, and whether activity in left hippocampus is modulated by event properties (i.e., the conditions), displaying the same linear trend as the behavioural responses. As this study is the first to investigate retrospective duration estimation using neuroimaging methods, a second aim of this study was to explore what other regions of the brain are involved in retrospective duration estimation.

In sum, given that left hippocampus plays a role in both thinking about events in the past, and in encoding temporal information, this area might play a role in retrospective duration estimation. As retrospective paradigms critically rely on recollecting events to judge their duration, left hippocampus and surrounding areas in the MTL are likely to be critical to this process (MacDonald, 2014).

5.3.2 Research hypotheses

As outlined above, given its involvement in both representing sequential event information as well as temporal anchoring, hippocampus is likely to play a role in retrospective duration estimation. Therefore, the first hypothesis was that hippocampus is involved in retrospective duration estimation. Secondly, the question was whether activity in hippocampus is modulated by event properties in

a similar fashion as the behavioural data presented in Chapter 2. As Chapter 2 has shown, retrospective duration ratings are modulated by the number of perceived segments and the perceived similarity between them, over and above clock duration. This means that retrospective ratings are based on the memory of the temporal unfolding of events. As argued above, in particular left hippocampus appears to be involved in representing this temporal unfolding in terms of event sequencing and temporal anchoring, and in representing event content in terms of retrieving episodic memories and binding event components. The hypothesis was that more perceived sub-events and less similarity between them should lead to more activity in left hippocampus during retrospective duration estimation. Furthermore, we explored what other regions are involved in duration reconstruction. The same questions were asked for the recognition memory task, providing an overview in regions involved and investigating whether activity in left hippocampus increased over conditions as an effect of the amount of event detail that needs to be recollected.

5.4 Methods

5.4.1 Participants

Thirty-seven right-handed native English speakers were recruited from the University of York student population. Participants received course credit or a small monetary reward as compensation. One participant was excluded due to poor memory performance (false alarm rate <50%). All other participants displayed correct overall recognition of the probes over 50% ($M=85\%$, $SD=9.6\%$). There was no difference between conditions as per design (Friedman's test n.s., basic: $M=90\%$, numerous: $M=83\%$, dissimilar: $M=82\%$). Participants had normal or corrected-to-normal vision. Ethical approval for this experiment was obtained from the York Neuroimaging Centre Research Ethics Committee.

5.4.2 Materials

The same materials used in Chapter 2, Experiment 2 were used in this experiment.

5.4.3 Task design

The same tasks used in Chapter 2, Experiment 2 were used in this experiment. The experiment consisted of a study phase, a recognition memory task and a duration estimation task. Participants performed the study phase outside of the fMRI scanner. After studying all animations three times, they were asked whether they were confident that they had remembered all animations. If they did not feel confident, they performed a fourth study cycle.

After the study phase, participants performed a response practice task to ensure that they were familiar with the button responses that they were to use in the scanner. As two fMRI-compatible button boxes with two buttons each were used to collect responses in the scanner, participants practiced responding using these four buttons. For the recognition memory task, participants were shown probes and foils belonging to the anchor animations and were asked to respond with their right index finger when they believed that the frame belonged to one of the animations that they have studied, and to use their left index finger when they believed it was a visually similar foil. Participants then performed a practice task for indicating the numbers 1 to 7 using their left and right index and middle fingers. These numbers were mapped onto the four buttons in a way that one press of the left middle finger signalled '1', left index finger '2', right index finger '3', right middle finger '4', two presses of the left middle finger '5', two presses of the left index finger '6' and two presses of the right index finger '7'. They were told that they would have to provide a rating on a scale of 1 to 7 in the scanner, but not what they would be rating. Experimenters ensured that participants were completely confident with both types of responses before moving the participant to the fMRI scanner.

Four fMRI scans were conducted: two scans were obtained while participants performed the recognition memory task and two scans while they

performed the duration rating task. Similar to Experiment 2, participants always performed the recognition memory task before the duration rating task. Both tasks were presented as a self-paced event related design: as soon as a participant responded to a stimulus, the paradigm moved on to the next stimulus or inter-stimulus interval. The optimal sequential ordering of events (conditions) and the duration of the inter-stimulus intervals was determined using Optseq2 software (<http://surfer.nmr.mgh.harvard.edu/optseq/>). In the recognition memory task, stimuli were either followed by the next stimulus, or by an inter-stimulus interval. In the duration estimation task, stimuli were always followed by an inter-stimulus interval. Stimuli were presented and timings were collected using Presentation version 12.2 (<http://www.neurobs.com>) and responses were obtained using fMRI compatible button boxes.

5.4.4 Acquisition parameters

Imaging was performed on a 3T GE Signa Excite MRI scanner at the York Neuroimaging Centre (YNiC). High-resolution whole brain T1-weighted structural images were obtained for all participants (1mm x 1mm x 1mm), and a T1-weighted FLAIR image was obtained to aid co-registration. Functional images were obtained using a gradient echo EPI sequence (TR=2 seconds, TE=50 ms, flip angle=90°, matrix=64x64, field of view=24 cm) with 3.5 mm thick axial slices.

5.4.5 Data analysis

fMRI data processing was carried out using FSL Version 5.98 (the software library of the Oxford Centre for Functional MRI of the Brain (FMRIB); www.fmrib.ox.ac.uk/fsl). Preprocessing of the data included BET brain extraction to remove scalp tissue and improve co-registration, motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), slice-timing correction using Fourier-space time-series phase-shifting, spatial smoothing using a Gaussian kernel of FWHM 9.0 mm, grand-mean intensity normalisation of the entire 4D dataset and high-pass temporal filtering (sigma=50.0s). Registration to high-resolution

structural and standard space images was carried out using FLIRT. First- and higher-level analyses were carried out using FEAT (FMRI Expert Analysis Tool). Time-series statistical analysis was carried out using FILM (FMRIB's Improved Linear Model) with local autocorrelation correction (Woolrich, Ripley, Brady, & Smith, 2001).

Both tasks were modelled separately, and both sessions of each task were modelled separately (only the first session of each task was used for further analysis because there was sufficient power to detect differences and so there was no need to aggregate the sessions). Given the self-paced event-related design of the tasks, events were modelled for each individual participant, with the onset for each event being the onset of the stimulus presentation, and the duration being the time at which the participant responds. This aimed to capture the recognition decision and duration reconstruction process for each item for each individual participant. Each event condition was modelled as a separate explanatory variable. For each individual participant, the contrasts for the three conditions relative to baseline (basic: 1, 0, 0; numerous: 0, 1, 0; and dissimilar: 0, 0, 1) and the linear contrast (basic < dissimilar: -1, 0, 1) were computed, and Z-statistic images were thresholded at $Z=1.96$, ($p=.05$, uncorrected for multiple comparisons).

High-level analyses were carried out using FLAME. Given that this was the first study to use a retrospective duration estimation paradigm in the fMRI scanner, it was not known from previous research what other regions might be involved in this process. Therefore, an exploratory high-level group analysis was also conducted. The effect of all conditions over baseline (1, 1, 1) was investigated for the duration rating task, yielding an overview of areas involved. To investigate the network of regions involved in the recognition task, the effect of all conditions over baseline (1, 1, 1) was investigated. For both analyses, Z-statistic images were thresholded using clusters determined at $Z>2.3$ and a (corrected) cluster significance threshold of $p=.05$.

5.4.5.1 Region of interest analysis

Left hippocampus is considered to play a central role in retrieving event information, and activity in this region has been shown to increase when more event information needs to be retrieved (Addis & Schacter, 2008). To test whether activity in the left hippocampus was modulated by the conditions during duration reconstruction and recognition, an atlas-defined region of interest (ROI) of the left hippocampus (Harvard-Oxford Subcortical Structural Atlas) was applied at the high-level analysis and interrogated for the linear contrast (-1, 0, 1; basic < dissimilar). Activity in the ROI was corrected using Gaussian Random Field theory-based maximum height thresholding with a corrected significance threshold of $p=.05$. In order to visually inspect whether the activity followed the same pattern as the behavioural data, a 9.0 mm sphere around the maximum was used to extract the individual percent signal change using FEATquery.

5.4.5.2 Second-level modulation of event properties

Furthermore, to investigate what areas of the brain are involved in retrieving event properties, an exploratory parametric analysis was conducted using the ratings of the number of segments and the similarity between them (as obtained in Experiment 2) as covariates. This was done by demeaning the scores for the number of segments and the similarity between them and adding them to the model design as the 'strength' or 'height' of the event. This additional parameter was used to modulate the height (magnitude) of the predicted response, and thus yielded regions of the brain in which the response followed the pattern of the ratings. Both the number of segments and the similarity between them were compared to baseline using the contrast (1 0) and (0 -1) respectively (-1 as we are interested in areas that respond to a decrease in similarity), providing an overview of brain areas in which activity increased with the number of segments, and regions where activity increased with a decrease in similarity. Furthermore, a conjunction analysis was conducted to investigate in what regions activity increased as an effect of an increase in the number of segments and a decrease in similarity (1 -1). As this was an exploratory analysis that might generate

hypotheses for further research, a very liberal threshold was used ($Z > 2.3$, which corresponds to $p = .01$ one-sided, uncorrected for multiple comparisons). In analogy with the behavioural regression analyses, we also investigated what regions correlated with actual duration, following the same procedure with clock duration as a covariate.

5.5 Behavioural results

5.5.1 Recognition memory

Repeated measures ANOVAs were used to investigate the effect of condition on the response latencies in the recognition memory task. Repeated measures ANOVAs with response latencies for only the correctly recognised probes aggregated up to the subject (F_1) or item (F_2) level as a dependent variable did not reveal a significant main effect of condition on the response latencies ($F_1(2, 70) = .92, p = .40, \eta^2 = .03$; $F_2(2, 54) = 2.37, p = .10, \eta^2 = .10$), and no significant linear trends (although marginal when aggregated to the item level) ($F_1(1, 74) = 1.84, p = .18, \eta^2 = .05$, $F_2(1, 27) = 3.72, p = .06, \eta^2 = .12$).

Due to practical limitations, the number of participants that could be scanned for this fMRI study was lower than in the behavioural counterpart of this experiment (36 here compared with 75 in Experiment 2, Chapter 2). Hence, the present study lacked power compared with Experiment 2. Furthermore, statistical methods to analyse event-related fMRI data often require high statistical power, normally obtained by increasing the number of items or repetitions of items. Given that this was not an option for the current design (the number of items that participants can remember is limited) the fMRI data presented here was based on both correct and incorrect probe responses to keep the number of data points as large as possible (note that more fMRI data will be acquired for this study when practical circumstances allow). Therefore, the present behavioural data was reanalysed for both correctly and incorrectly recognised probes to increase

statistical power and to obtain a prediction for the pattern of brain activity in this task (i.e., whether a linear trend could be expected based on behaviour).

Taking both correct and incorrect responses probes into account, the results of repeated measures ANOVAs with response latencies aggregated up to the subject (F_1) or item (F_2) level as a dependent variable indicated that there was a significant main effect of condition on the response latencies ($F_1(2, 70)= 2.45$, $p=.09$, $\eta =.07$; $F_2(2, 54)=7.26$, $p=.002$, $\eta =.21$), although marginal over subjects, and significant linear trends ($F_1(1, 74)= 4.24$ $p=.047$, $\eta = .11$, $F_2(1, 27)=11.07$, $p= .003$, $\eta =.29$) (Figure 21). Although based on both correct and incorrect responses, these findings are consistent with the patterns found in Experiment 2, Chapter 2, suggesting that properties of the events encoded in memory during the study phase affect the response latencies of memory judgements, with the conditions that require recollection of more event information leading to increased latencies.

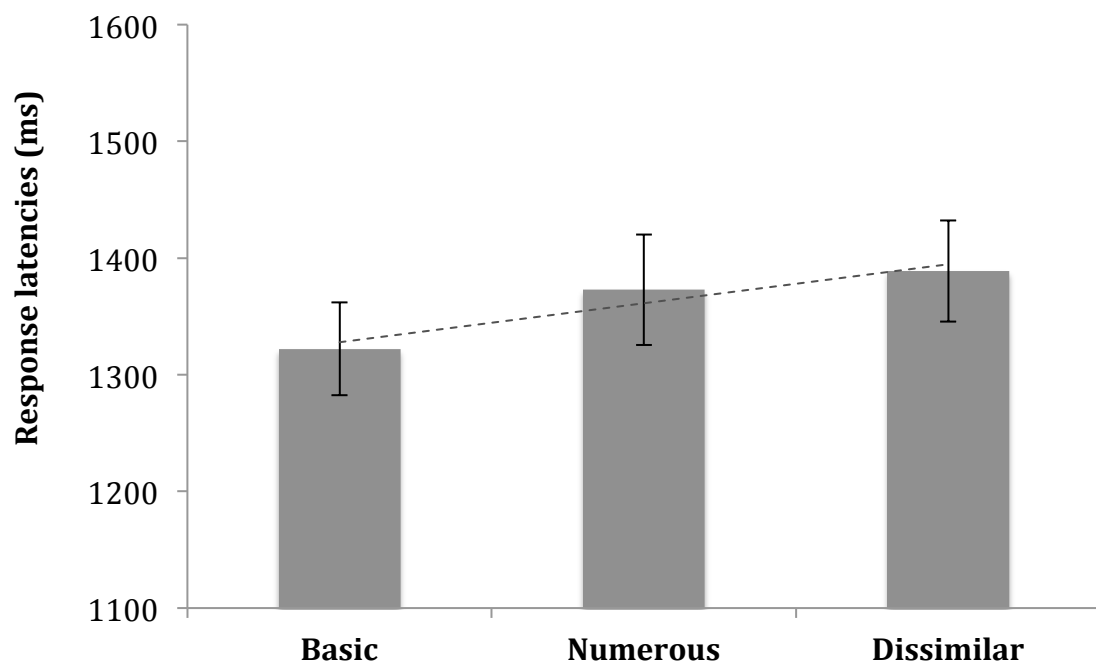


Figure 21. Response latencies over items. Error bars represent standard error. Trend line represents linear trend over conditions.

5.5.2 Duration ratings

Repeated measures ANOVAs were used to investigate the effect of condition on the duration ratings. The results of repeated measures ANOVAs with duration ratings aggregated up to the subject (F_1) or item (F_2) level as a dependent variable indicated that there was a main effect of condition on the duration ratings of the animations ($F_1(2, 66) = 8.79, p < .001, \eta_p^2 = .21$; $F_2(2, 54) = 14.14, p < .001, \eta_p^2 = .34$) taking items into account that were either correctly or incorrectly recognised in the probe recognition task. Furthermore, the results indicated significant linear trends ($F_1(1, 33) = 21.88, p < .001, \eta_p^2 = .40$; $F_2(1, 27) = 27.77, p < .001, \eta_p^2 = .51$) (Figure 22). This pattern of results remained the same if only correctly recognised items were taken into consideration ($F_1(2, 66) = 5.82, p = .005, \eta_p^2 = .15$; $F_2(2, 54) = 13.72, p < .001, \eta_p^2 = .34$; linear trends $F_1(1, 33) = 15.05, p < .001, \eta_p^2 = .31$; $F_2(1, 27) = 24.86, p < .001, \eta_p^2 = .51$). These findings are similar to the results from Experiment 2, Chapter 2, suggesting that the event properties encoded in memory are retrieved during the duration judgement task, with more event information leading to longer duration ratings.

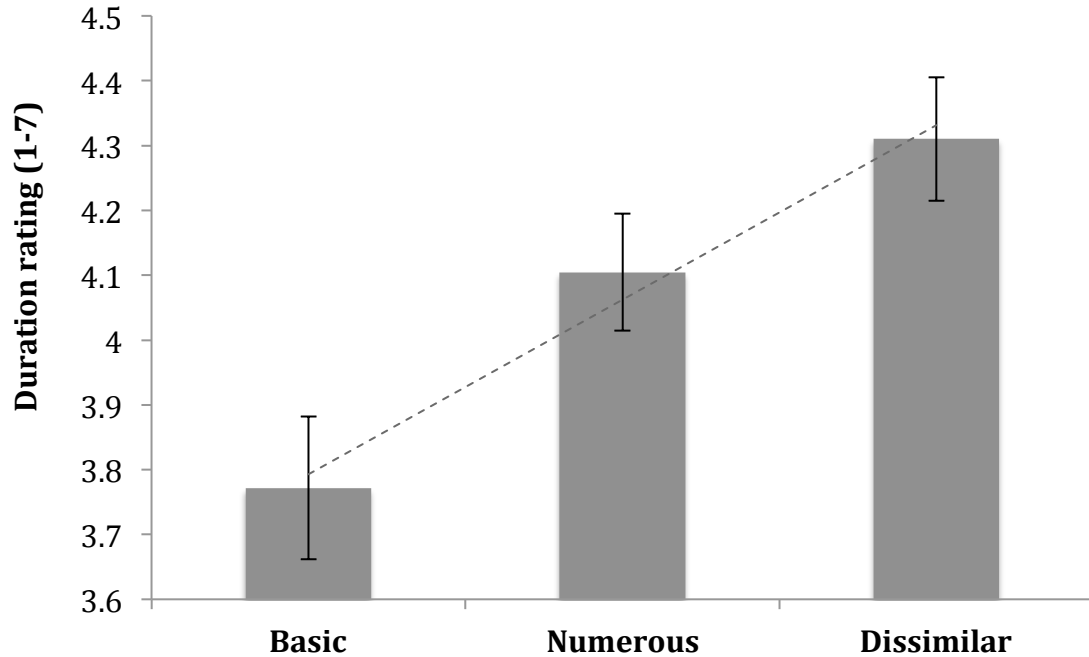


Figure 22. Duration ratings over subjects (scale 1-7). Error bars represent standard error. Trend line represents linear trend over conditions.

5.5.3 Regression results

Similar to the previous chapters, a by-item hierarchical multiple regression analysis was conducted to investigate the effect of event properties on the behavioural duration judgements. The proportion of variance accounted for by the number of sub-events and the similarity between them (as obtained in Experiment 2) over and above actual clock duration was investigated. Clock duration was added to the model first as a control predictor to account for the systematic variation in duration between the triads. The results show that adding the number of segments to the regression model increased the proportion of variance account for (R) significantly from .45 to .59 ($F_{\text{change}}(1, 81) = 18.76, p < .001$). Adding the dissimilarity between the segments to this regression model again increased the proportion of variance account for, from .59 to .66 ($F_{\text{change}}(1, 82) = 11.39, p = .001$) (Table 8). Changing the order in which the number of segments and similarity between events were added to the model did not affect the pattern of significance.

Table 8. Regression coefficients for duration ratings. Note: ** indicates $p \leq .001$

	Model	<i>B</i>	<i>SE B</i>	β
1.	Constant	2.917	.252	
	Clock duration	.18	.04	.45**
2.	Constant	2.699	.23	
	Clock duration	.04	.05	.10
	Number of sub-events	.23	.05	.52**
3.	Constant	3.672	.36	
	Clock duration	.02	.05	.06
	Number of sub-events	.23	.05	.52**
	Similarity	-.19	.06	-.29**

5.6 fMRI results

5.6.1 Recognition memory

5.6.1.1 Whole brain results recognition memory

To investigate network of regions involved in probe recognition from memory, a whole brain analysis was conducted looking at the effect of probe recognition over baseline (contrast: 1, 1, 1) (cluster-corrected, $Z > 2.3$, $p < .05$). This analysis revealed high levels of activity in visual and motor areas ($Z > 10$), as expected by the nature of the task (i.e., visual stimulation and motor responses). Furthermore, high levels of activity were found in bilateral inferior frontal gyrus, insula, thalamus, hippocampus, putamen, and right cerebellum ($Z > 8$) (Figure 23; Table 9). As expected, left hippocampus was indeed involved in the recognition memory task.

Table 9. Recognition memory over baseline. Local maxima are reported ($Z > 8$, local maxima $> 9\text{mm}$ apart).

Brain area	Hemisphere	Z	Voxel (MNI)		
			x	y	z
Primary visual cortex	left	14.6	-2	-86	-6
Ventral occipital cortex	left	13.7	-20	-78	-14
Lateral occipital cortex	right	13.6	26	-66	42
Lateral occipital cortex	right	13.4	32	-80	14
Lateral occipital cortex	left	13.3	-28	-88	6
Lateral occipital cortex	right	13.1	30	-70	28
Lateral occipital cortex	left	13.1	-22	-68	38
Anterior cingulate cortex	left	13.0	-2	8	46
Supramarginal gyrus	left	12.8	-30	-56	38
Precentral gyrus	right	12.6	46	4	26
Insula	right	12.5	32	22	-4
Insula	left	12.1	-30	20	-2
Precentral gyrus	left	11.9	-44	0	26
Precentral gyrus	left	11.8	-26	-8	48
Precentral gyrus	right	11.7	30	0	48
Inferior frontal gyrus, pars triangularis	right	11.6	42	30	14
Inferior frontal gyrus, pars opercularis	left	10.5	-40	22	18
Hippocampus	right	9.68	20	-32	-8
Hippocampus	left	9.28	-20	-34	-8
Thalamus	left	8.78	-14	-22	4
Thalamus	right	8.66	12	-18	4
Putamen	right	8.64	16	8	-2
Cerebellum	right	8.51	26	-42	-48
Putamen	left	8.11	-14	4	-2

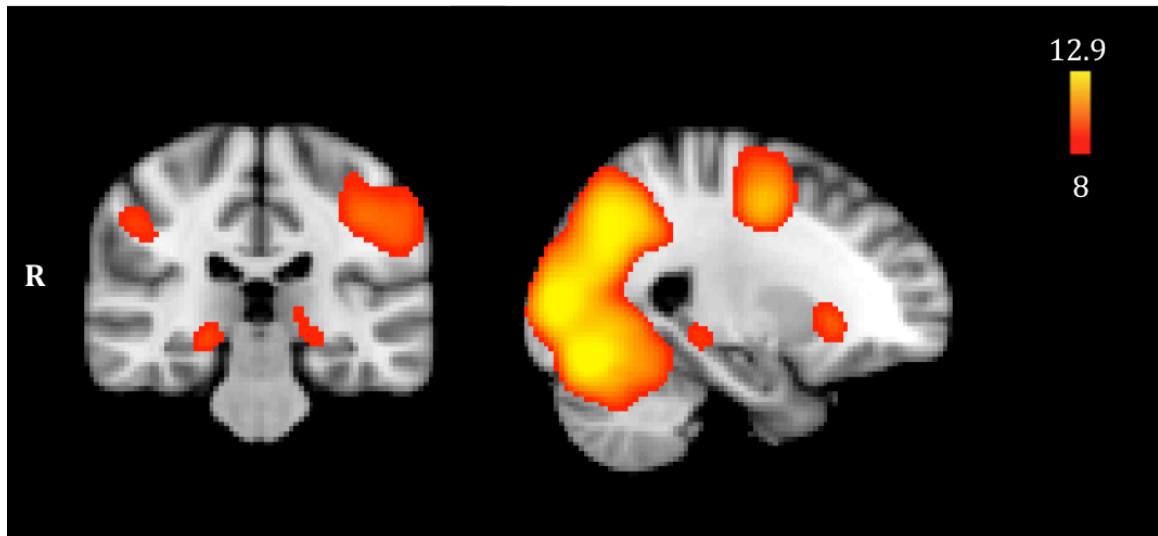


Figure 23. Whole brain results for probe recognition over baseline (contrast: 1, 1, 1). Coordinates: -24, -28, -10 in MNI space. Left panel shows activity in left and right hippocampus. Right panel shows activity in left hippocampus.

5.6.1.2 Region of interest results recognition memory

To investigate whether activity in the hippocampus for the recognition memory task was modulated by event structure as conveyed by the difference between the basic and dissimilar conditions (i.e., the linear trend), an ROI analysis was conducted using a mask of left hippocampus as defined by the Harvard-Oxford Subcortical Atlas. The results showed that within the hippocampus ROI, there was no activity for the dissimilar condition over the basic condition on the statistical threshold of $p < .05$ corrected. Even when a very liberal uncorrected threshold was used ($p < .05$ uncorrected), no activity was observed in hippocampus for the dissimilar over the basic condition. These findings suggested that there is no effect of condition in the shape of a linear trend in left hippocampus for the recognition memory task.

To explore whether any other regions showed a linear trend over conditions, an exploratory whole brain analysis was conducted for the linear contrast. At the very liberal threshold of $p < .05$, uncorrected for multiple comparisons, one cluster was observed, including right inferior parietal lobule (cluster size: 162 voxels, $Z = 2.72$, coordinates: 52, -70, 28 in MNI space). This

structure has previously been implicated in recognition memory tasks, and in particular in old/new discrimination tasks (Wagner et al., 2005). However, given that the present data contained both correctly and incorrectly recognised items, we were limited in our interpretation of the effect of condition on activity in this region.

5.6.2 Duration reconstruction

5.6.2.1 Whole brain results duration reconstruction

To investigate the network of regions involved in retrospective duration reconstruction, a whole brain analysis was conducted looking at the effect of duration reconstruction over baseline (contrast: 1, 1, 1) (cluster-corrected, $Z > 2.3$, $p < .05$). This analysis revealed high levels of activity in visual and motor areas ($Z > 13$), as expected by the nature of the task (i.e., visual stimulation and motor responses). Furthermore, high levels of activity were found in bilateral inferior frontal gyrus, insula, thalamus, hippocampus, putamen and cerebellum ($Z > 8$) (Figure 24, Table 10). These results suggest that both duration reconstruction and recognition memory recruit the same network of regions, and that left hippocampus is indeed involved in reconstructing duration.

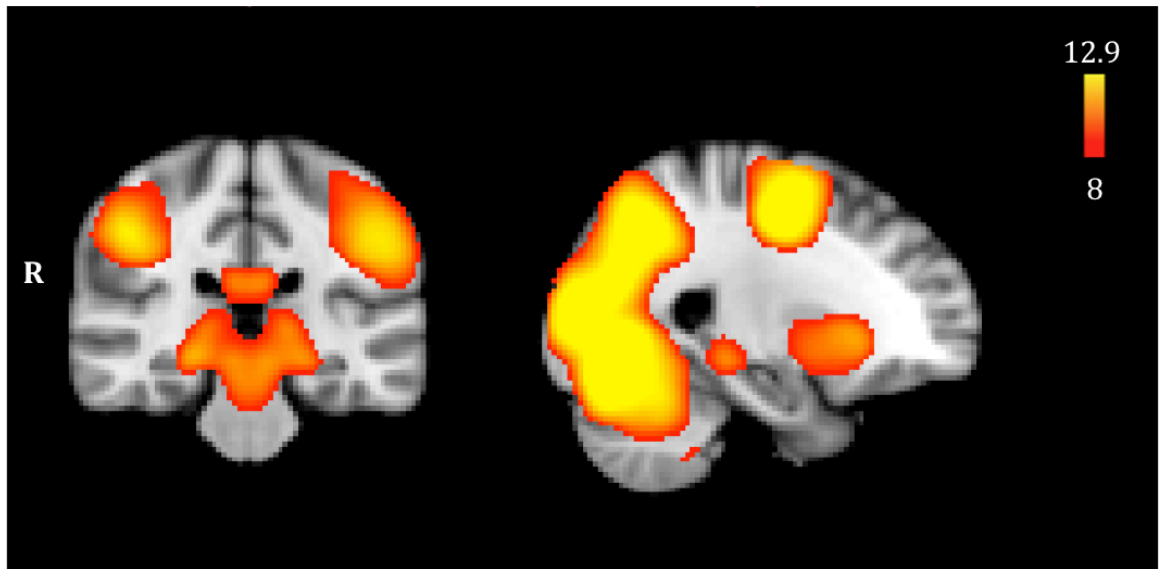


Figure 24. Whole brain results for duration reconstruction over baseline (contrast: 1, 1, 1). Coordinates: -24, -28, -10 in MNI space. Left panel shows activity in left and right hippocampus. Right panel shows activity in left hippocampus.

Table 10. Duration reconstruction over baseline. Local maxima are reported ($Z > 8$, local maxima $> 9\text{mm}$ apart).

Brain area	Hemisphere	Z	Voxel (MNI)		
			x	y	z
Primary visual cortex	right	16.5	6	-84	0
Anterior cingulate gyrus	left	15.4	-4	8	44
Lateral occipital cortex	left	14.5	-28	-88	8
Lateral occipital cortex	right	14.3	34	-80	16
Precentral gyrus	left	14.3	-26	-8	50
Supramarginal gyrus	left	14.1	-42	-38	36
Parietal lobule	left	14.1	-30	-56	42
Precuneus	left	13.9	-20	-70	44
Precentral gyrus	right	13.8	28	-4	48
Supramarginal gyrus	right	13.8	24	-38	40
Precuneus	right	13.8	26	-66	46
Precentral gyrus	right	13.7	48	4	26
Insula	right	13.7	32	20	0
Insula	left	13.6	-32	16	2
Precentral gyrus	left	13.6	-46	-2	26
Inferior frontal gyrus, pars triangularis	right	12.6	44	32	16
Inferior frontal gyrus, pars triangularis	left	12.2	-38	24	18
Thalamus	right	11.9	12	-18	2
Thalamus	left	11.7	-12	-20	0
Hippocampus	right	11.6	22	-30	-8
Putamen	right	11.5	18	8	-2
Putamen	left	11.3	-18	6	-4
Hippocampus	left	11.2	-20	-32	-8
Posterior cingulate gyrus	right	10.7	-2	-30	22
Cerebellum	right	9.51	26	-42	-48
Cerebellum	left	8.72	-22	-44	-48

5.6.2.2 Region of interest results duration reconstruction

To investigate whether activity in the hippocampus during duration reconstruction was modulated by event structure as conveyed by the difference between the basic and dissimilar conditions (i.e., the linear trend), an ROI analysis was conducted using a mask of left hippocampus as defined by the Harvard-Oxford Subcortical Atlas. The results showed that within the hippocampus ROI, the dissimilar condition elicited significantly more activity in left hippocampus than the basic condition (contrast -1, 0, 1; $p < .05$ corrected), suggesting a linear trend over conditions (Figure 25). To further investigate whether neural activity in left hippocampus followed the same pattern as the behavioural data (i.e. basic \leq numerous \leq dissimilar), BOLD percent signal change was extracted for each condition across participants from a 9mm radius sphere that was drawn around the maximum of activity in the ROI (Figure 26). Visual inspection of the bar graphs presenting the BOLD percent signal change suggest that indeed, activity around the local maximum follows the same pattern as the behavioural data (Figure 27). These findings are in line with the hypothesis that activity in hippocampus should be modulated by condition, as more information needs to be reactivated to reconstruct the duration of events with more sub-events and less similarity between them.

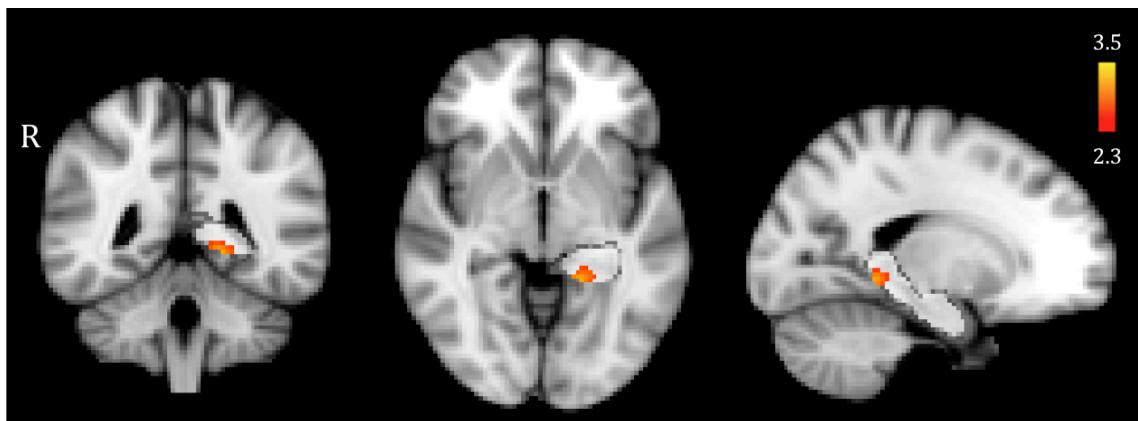


Figure 25. Contrast basic < dissimilar in left hippocampus for duration reconstruction ($p < .05$, corrected). The predefined region of interest is outlined in grey (MNI: -18, -44-4).

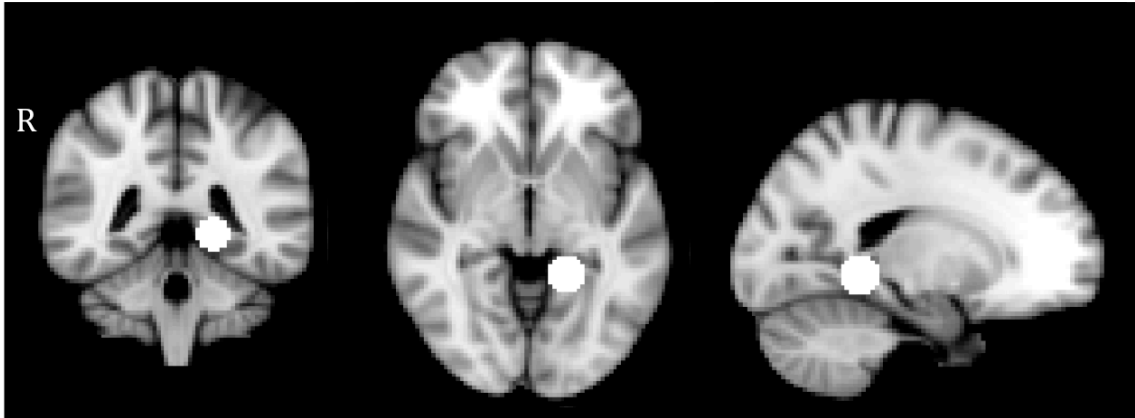


Figure 26. Spherical ROI (9mm radius) drawn around the local maximum within the left hippocampus (MNI: -18, -44-4).

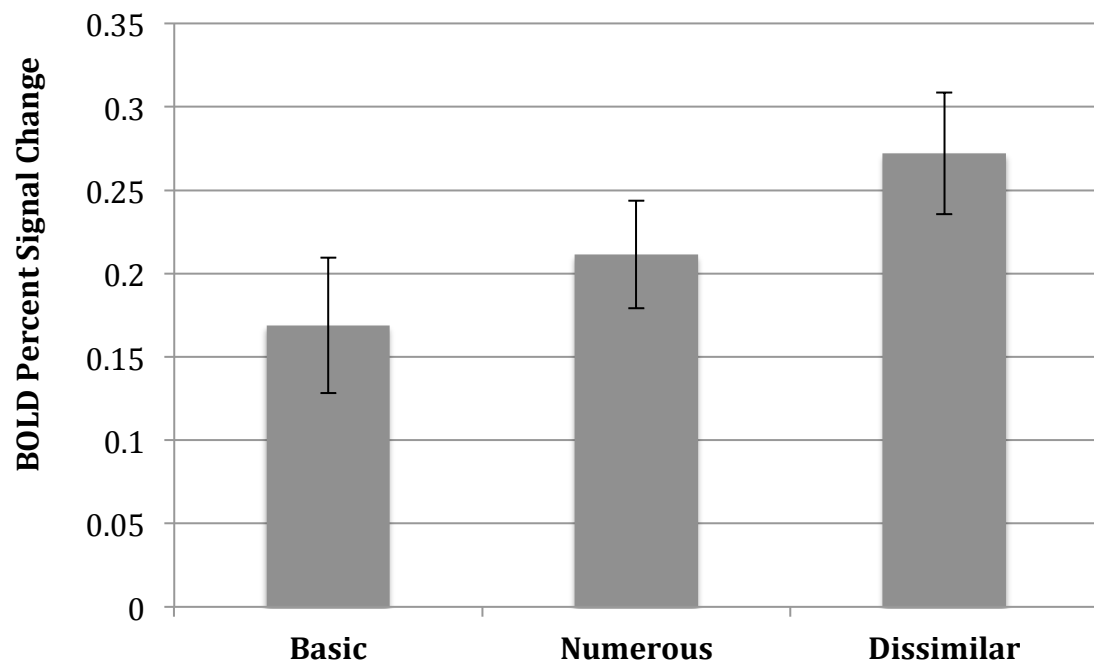


Figure 27. BOLD percent signal change extracted from the spherical ROI (figure 26) over participants per condition.

To explore whether any other regions showed a linear trend over conditions, an exploratory whole brain analysis was conducted for the linear contrast. At the very liberal threshold of $p < .01$ (one-sided) uncorrected for multiple comparisons, two cluster larger than 50 voxels were observed. As expected, a large cluster was observed in the left posterior hippocampus / lingual gyrus (cluster size: 359 voxels, $Z=3.16$, coordinates: -16, -44, -4 in MNI space; Figure 28). A small cluster was observed in right cerebellum (cluster size: 58 voxels, $Z=2.55$, coordinates: 34, -60, -26 in MNI space). However, as cerebellum has not been scanned consistently (i.e. not always within the field of view) in this experiment, these results need to be interpreted with care, as it is unclear which subjects contributed to this effect. Overall, the results from this exploratory analysis suggested that only a cluster in left hippocampus / lingual gyrus displayed the linear trend over conditions.

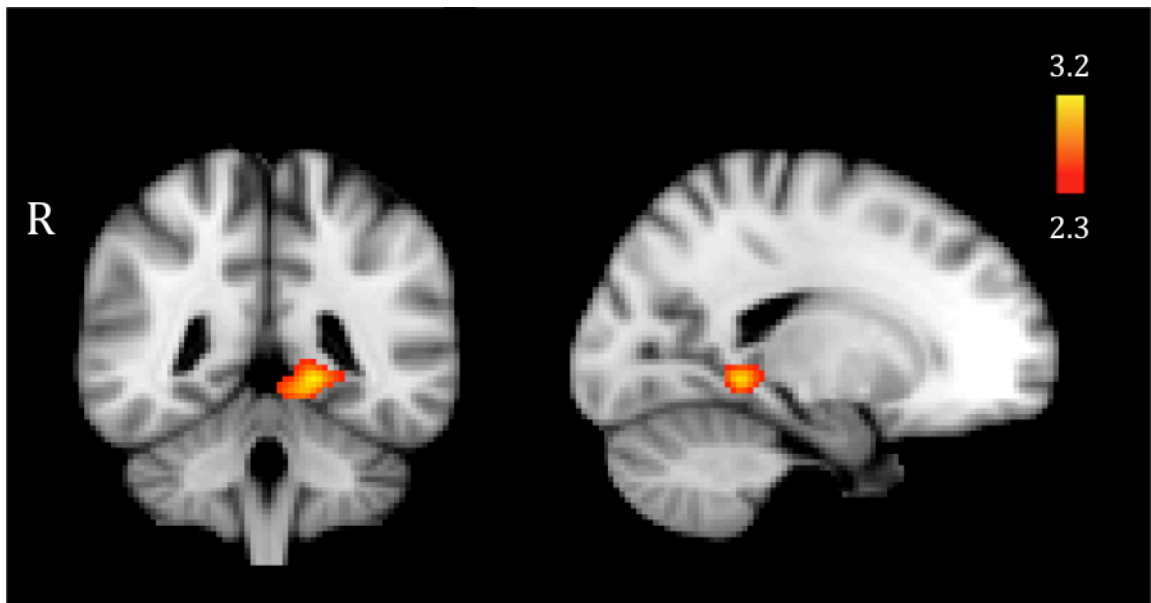


Figure 28. Contrast basic < dissimilar for duration reconstruction on a liberal threshold ($Z=2.3$, equals $p < .01$ one sided, uncorrected for multiple comparisons) (MNI: -18, -44-4).

5.6.2.3 Whole brain results duration reconstruction and event properties

To explore whether activity in areas of the brain increases as an effect of a higher number of segments or a higher level of dissimilarity between segments, a whole brain analysis was conducted looking at the contrasts of the number of segments versus baseline (1, 0) and the similarity versus baseline (0, 1). A very liberal threshold was used, as this analysis was exploratory in nature and may generate testable hypotheses for further studies ($Z > 2.3$, equals $p < .01$ one-sided, uncorrected for multiple comparisons).

For the number of segments, the analysis revealed a large cluster in the right amygdala, showing overlap with the cornu ammonis of the right hippocampus. A smaller cluster was found in the left amygdala and hippocampal region (see Figure 29, bottom panel). As expected, a cluster was found in the right lingual gyrus. Furthermore, clusters were found in early visual areas, cerebellum and the left frontal orbital cortex and insula (see Table 11).

For similarity, the whole brain analysis only revealed one cluster larger than 50 voxels, which was located in the early visual areas. A small cluster was found in the left lingual gyrus, which overlaps with activity found for the retrospective duration estimation task (see Figure 29, upper panel). Very small clusters were found in the right insula and right hippocampus (see Table 12).

As a control, we investigated what areas of the brain correlated with clock duration. We found no meaningful activity correlated with clock duration (i.e., only in white matter and no overlap with areas reported below).

Table 11. Whole brain results for number of segments ($Z > 2.3$, uncorrected). Coordinates represent peak voxels of uncorrected clusters. Area names obtained from Harvard-Oxford Cortical Structural Atlas and Juelich Histological Atlas, and represent the location of the peak and direction of the cluster spread. All clusters > 50 voxels and are organised by size (largest to smallest).

Brain area	Z	<i>Voxels</i>	Voxel (in MNI-space)		
			x	y	z
Left V1	4.09	1562	-10	-104	-10
Right amygdala, Hippocampus cornu ammonis	3.41	544	30	-6	-14
Right middle temporal gyrus, Angular gyrus	2.81	251	62	50	10
Cerebellum	3.02	223	-4	-60	-42
Right lingual gyrus	2.72	128	18	-62	-4
Left amygdala, Hippocampus cornu ammonis	2.48	87	-30	-6	-14
Left frontal orbital cortex, Insula	2.67	75	-32	20	-12

Table 12. Whole brain results for perceived similarity ($Z > 2.3$, uncorrected). Coordinates represent peak voxels of uncorrected clusters. Area names obtained from Harvard-Oxford Cortical Structural Atlas and Juelich Histological Atlas, and represent the location of the peak. All clusters > 20 voxels and are organised by size (largest to smallest).

Brain area	Z	<i>Voxels</i>	Voxel (in MNI-space)		
			x	y	z
Right V1	2.69	216	26	-70	12
Left lingual gyrus, Fusiform gyrus	2.55	42	-25	-64	-10
Right insula	2.48	28	40	-6	-2
Right hippocampus cornu ammonis	2.43	21	22	-36	8

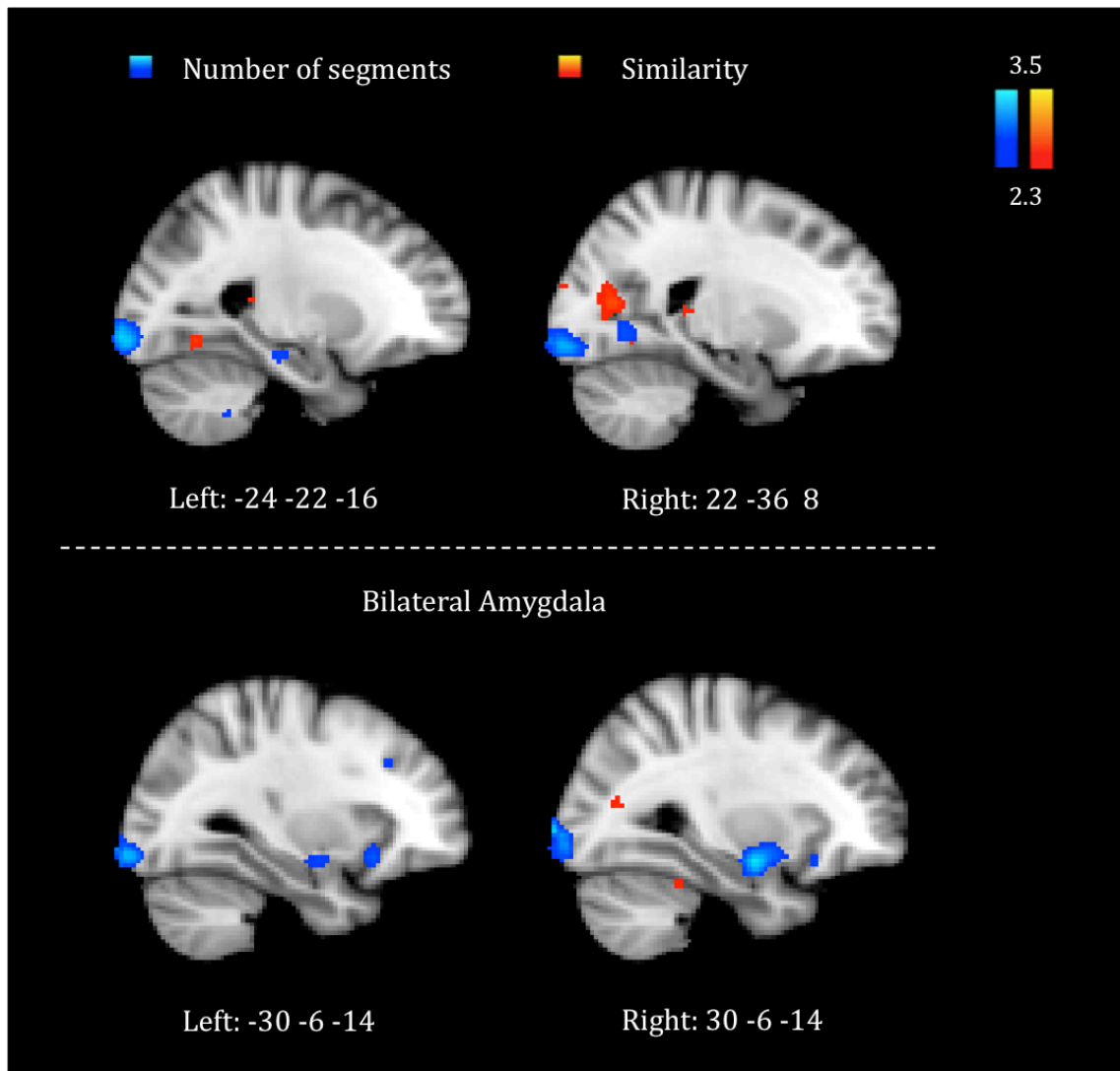


Figure 29. Whole brain results for number of segments and perceived similarity ($Z > 2.3$, uncorrected). Top panel: small clusters in the hippocampal region and lingual gyrus. Bottom panel: increase in bilateral amygdala activity for increase in number of segments.

Furthermore, a conjunction analysis was conducted to investigate which regions of the brain display activity that correlated with both an increase in number of segments and a decrease in similarity (1 -1). Similar to the analysis over conditions, the results include clusters in the right lingual gyrus and left and right hippocampus, suggesting that indeed activity in these regions might be modulated by event properties (Table 13, Figure 30). These findings are in line with the

prediction that hippocampus and visual areas such as the lingual gyrus play a role in reconstructing event representations. In sum, regions including lingual gyrus, hippocampus and the insular cortex might play a role in both the retrieval of the number of segments as well as the similarity between them.

Table 13. Whole brain results for an increase in the number of segments and a decrease in similarity (contrast: 1, -1) ($Z > 2.3$, uncorrected). Coordinates represent peak voxels of uncorrected clusters. Area names obtained from Harvard-Oxford Cortical Structural Atlas and Juelich Histological Atlas, and represent the location of the peak and direction of the cluster spread. All clusters > 50 voxels and are organised by size (largest to smallest).

Brain area	Z	Voxels	Voxel (in MNI-space)		
			x	y	z
Right lingual gyrus	3.4	2533	18	-62	-6
Right insula	3.13	1271	40	0	-6
Left inferior frontal gyrus	3.34	205	-54	30	-14
Left middle temporal gyrus	2.87	200	-52	-30	-2
Right hippocampus (subiculum)	2.89	85	26	-28	-10
Left hippocampus (subiculum), Parahippocampal gyrus	2.6	61	-22	-20	-28
Left frontal medial cortex	2.72	53	-2	44	-14

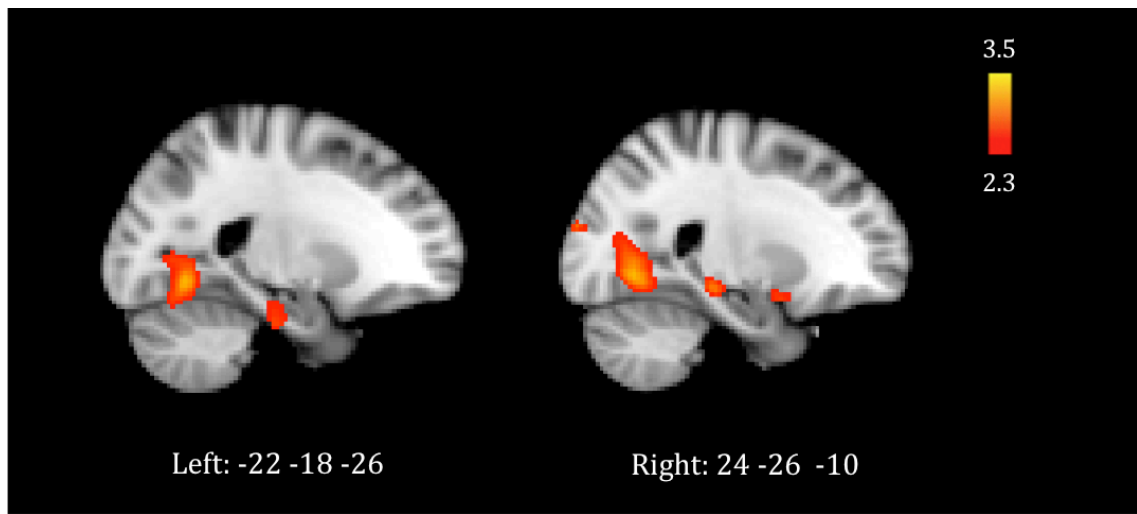


Figure 30. Whole brain results for an increase in the number of segments and a decrease in similarity (contrast: 1 -1) ($Z > 2.3$, uncorrected). Clusters in the left and right hippocampal region and lingual gyrus.

5.7 Discussion

The aims of the present study were to investigate whether hippocampus indeed plays a role in retrospective duration estimation, whether activity in hippocampus shows the same pattern as the behavioural results and what other regions may be involved in retrospective duration estimation. The results presented here confirm that hippocampus is involved in retrospective duration estimation, and that activity in this area is modulated by the conditions, as shown by the region of interest analysis. Furthermore, no positive trend in hippocampal activity was observed over conditions for the recognition memory task, suggesting that although both tasks recruit hippocampus, its sensitivity to event structure is only observed in tasks that require the retrieval of detailed event information. The fact that for the duration reconstruction task, no other regions responded (more) strongly to the linear contrast suggests that indeed, left hippocampus, with its sensitivity to event structure, plays a crucial role in duration reconstruction, with more retrieved event information leading to longer estimates.

A preliminary investigation analysing which areas of the brain show activity that increases with the number of segments and lower similarity between them reveal that activity in bilateral amygdala appear to increase with a higher number of segments. Furthermore, regions including lingual gyrus, hippocampus and the insular cortex might play a role in both the retrieval of the number of segments as well as the similarity between them, although the current results do not provide enough statistical evidence to allow for any firm conclusions. Furthermore, the behavioural results presented here provide yet another replication of the effect observed in Chapter 2: retrospective duration estimates display a linear trend over conditions, and both number of segments and similarity between them significantly affect these estimates, over and above actual duration.

5.7.1 Hippocampus and duration estimation

These findings contribute to the recent literature on the function of hippocampus in time and memory by providing evidence for the role of hippocampus in retrospective duration estimation. The present study extends previous findings by showing that hippocampus not only plays a role during the encoding of temporal order and anchoring, as shown by many of the rodent and a few human studies, but also during the *retrieval* of temporal information about events. Previous studies have not attempted to or succeeded to address retrospective duration estimation, and as such, this study provides the first evidence for the involvement of hippocampus in this process.

Even though studies up until now have not shown this involvement before, it was recently hypothesised by MacDonald (2014) who argued that hippocampus and surrounding areas in the MTL are likely to be critical to this process, as retrospective paradigms critically rely on recollecting events to judge their duration. Indeed, the current findings corroborate this. Moreover, the present findings suggest that activity in hippocampus during duration estimation is guided by event properties, and thus, dependent on the *content* of the interval being judged. This is in line with and extends findings from previous studies that suggest

that hippocampus plays a role in representing information about the temporal order of (sub-)events and their temporal anchoring, as the present study illustrates that indeed, event structure is employed by hippocampus during duration estimation.

In particular, the present study provides the first tentative evidence for a direct relationship between the retrieval of event properties and activity in MTL and other areas during duration estimation, as illustrated by the analysis of the effect of an increase in number of sub-events and decrease in similarity on activity on a whole brain level. Although this analysis itself does not provide sufficient statistical evidence for any firm conclusions, future studies could further focus on these regions using a region of interest approach. Nevertheless, there appears to be substantial overlap between the network revealed by the analysis of the duration estimation data over conditions and the network revealed by analysis of the data when not taking conditions into account but instead remodelling the data with the number of segments and similarity ratings as covariates. In particular, bilateral hippocampus, lingual gyrus and amygdala appear to be part of this network. The following will give a speculative account as to why this network of regions might be of interest to further research.

5.7.2 Event memory circuit and the role of hippocampus, lingual gyrus and amygdala

In order to reconstruct (the duration of) an event from memory, one needs to have a representation of the event content, information about the relevant event boundaries, and these representations need to be anchored in time and space. As argued above, representations of events consist of multimodal information about the content of events: for instance entities and their properties (e.g. colour, size) and information about their motion and trajectory can be part of an event representation. This visual information is thus likely processed by a subsystem of visual areas. Previous studies have found a network of posterior visual regions that is involved in remembering past events that includes the lingual, occipital and

fusiform gyrus (Addis et al., 2007). The regions that the present study found to be associated with reconstructing duration based on event properties are consistent with these findings. The fact that these regions appear to play a role in remembering or reconstructing past events is in line with the idea that sensory and perceptual areas that originally processed the input are reactivated upon retrieval (Buckner & Wheeler, 2001; Slotnick & Schacter, 2006; Addis et al., 2007). Thus, it appears to be the case that reconstructing the content of events to some extent recruits the same neural regions during encoding and retrieval, suggesting that these visual areas are part of a network of areas involved in remembering event information.

Furthermore, these regions in the lingual, occipital and fusiform gyri have been shown to be part of a functional network that also includes the posterior portion of the parahippocampal cortex and the hippocampus (Addis et al., 2007; Kahn et al., 2008). This functional network is one of two cortical pathways that have been shown to converge in the hippocampus (Kahn et al., 2008). The other pathway contains the anterior temporal lobe, middle temporal gyrus in the vicinity of the superior temporal sulcus converges via perirhinal and entorhinal cortex on the head of the hippocampus. These two pathways have been shown to be functionally distinct: they differ in terms of their functional connectivity as estimated using seed-based correlations during resting-state, and both pathways have been linked to different cognitive processes. Although many tasks activate both networks, the first pathway has been implicated in tasks such as recollection, autobiographical memory and event (re)construction (and as part of the default mode network), whereas the latter has been implicated in lexical and semantic processing, mentalising, face and voice recognition and also autobiographical memory (as summarised in: Kahn et al., 2008).

Hence, the novel finding that retrospective duration estimation recruits a network including (posterior) hippocampus, surrounding MTL regions and visual areas such as the lingual, occipital and fusiform gyrus is in line with its role in recollecting and (re)constructing events (Buckner & Wheeler, 2001; Slotnick & Schacter, 2006; Addis et al., 2007). The data moreover tentatively suggest that

activity in these regions is affected by event properties such as the number of events and the similarity between them – analogous to the results from the regression analysis performed on the behavioural rating data. Although not exhaustively demonstrated, the present results suggest that the information that is encoded and retrieved from these visual areas during event reconstruction can be described in terms of event characteristics, pointing toward more general underlying principles of information extraction and human memory (e.g. statistical learning).

However, this does not imply that the number of event boundaries and similarity between sub-events affect the same regions (equally). Based on the findings from the regression analysis of the behavioural data that show that the number of segments and similarity do not explain the same variance (i.e. they explain variance over and above each other), one might expect a similar differentiation on the neural level. The separate analyses of both the number of segments and similarity suggest that the regions overlap to some extent, but also that firstly, the number of segments modulates activity in more regions and secondly, that some regions appear to be modulated more strongly by the number of segments than by similarity. Although it is not possible to fully interpret these similarities and differences because of statistical limitations (both related to this design and to the novelty of this study and hence, the lack predefined regions of interest), there appears to be activity in the amygdala that is modulated by the number of segments. This is in line with findings from previous studies suggesting a role for amygdala in sequence learning (Schendan, Searl, Melrose, & Stern, 2003).

A very recent study by Willems and colleagues (in press) shows similar activity in right amygdala when participants listen to narratives. This activity is modulated by the mutual information of words: when the surprisal of hearing a word in the context of its predecessors is higher, more activity occurs in the right amygdala. In essence, this is a measure of prediction error, as amygdala becomes more activated when the upcoming word ‘violates’ the prediction of what word should come next. As argued in Chapter 1 (Figure 2), the perception of event boundaries also relies on monitoring prediction error: when there is a transient

rise in error rate, an event boundary is perceived and the event model is updated. Hence, the number of perceived event boundaries might be a memory counterpart of prediction error, as the number of perceived event boundaries is related to the number of times that a prediction has been violated. Amygdala may thus play a critical role in monitoring prediction error and/or updating event models. Although the link between the findings from the present study and the study by Willems and colleagues is speculative, further research could investigate what the function of amygdala is in event segmentation and memory for event boundaries.

Furthermore, as shown by the prospective studies in Chapter 4, the number of segments affects retrospective as well as purely prospective duration estimates, suggesting that the number of segments also plays a role – to some extent – when encoding duration (as opposed to content). The present study provides tentative evidence that temporal encoding in the hippocampus may be guided by event properties, albeit based on the assumption that areas that were active during encoding are reactivated during retrieval. This observation motivates further research into the role of MTL in encoding and retrieving information about time and event boundaries. In particular, as the present study did not investigate the encoding process, further studies could examine whether temporal encoding in the hippocampus is biased by event properties, and to what extent memory representations of content and duration overlap.

As argued above, obtaining a memory representation of event boundaries and similarity relies on monitoring event boundaries (i.e. monitoring prediction error), storing this information, and extracting information about regularities over subsequent exposures. Indeed, previous studies have suggested that temporal regularity (or in other words, transition probability) between objects might play a role in how event representations arise (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013). Schapiro and colleagues (2014) have shown that patients with hippocampal damage show gravely impaired statistical learning, and argue that detecting and learning regularities may be the domain of MTL rather cortex. The present study thus provides further evidence for a role for hippocampus in learning about event structures through statistical learning, showing that the

involvement of MTL in duration estimation may be due to reactivation of this knowledge.

Moreover, the ability of hippocampus to rapidly encode relations between arbitrary elements of sensory information is crucial to both statistical learning and representing event information (Schapiro et al., 2014). Hence, “hippocampus is unique in its ability to integrate and take into account the successive events and the context in which they appeared, and thus it is capable of creating such branching representations” (Paz et al., 2010, p. 6048). The findings of the present study, as supported by these properties of hippocampus, suggest that indeed, the involvement of hippocampus in representing “time” may be due to its role in building these associative representations of events, their relationships and their temporal unfolding through statistical learning.

5.7.3 Event perception in the brain

The present study sheds light on the question what information is retrieved when estimating the duration of novel events. The findings suggest that the number of event boundaries and the similarity between them modulates both behavioural estimates on an arbitrary scale and activity in the hippocampus. As pointed out above, it is currently unclear what the relationship is between monitoring and perceiving event boundaries and retrieving event information for the purpose of duration estimation. Some research has been done to investigate what areas of the brain are involved in monitoring upcoming event boundaries and updating event models (see Zacks et al., 2007 for an overview). In these studies, participants passively viewed dynamic movies of human actions (Zacks, Braver, et al., 2001) or simple two-object animations (Zacks, Swallow, Vettel, & McAvoy, 2006), or they read narratives (Speer, Reynolds, Swallow, & Zacks, 2009). For the studies of visual events, regions in extrastriate visual cortex including temporal, occipital and parietal areas displayed activity that correlated with the participant’s later identification of event boundaries. The activity in extrastriate visual areas observed in the current study is in line with these findings, suggesting that the

activity observed in these regions may be due to reactivation of observed event boundaries.

Furthermore, previous studies revealed that activity in parahippocampal areas transiently increased as an effect of spatiotemporal discontinuities in movies, suggesting that these areas play a role in integrating spatiotemporal event information (Magliano & Zacks, 2011). This increase was also observed in the bilateral insula and lateral-posterior temporal lobes. Discontinuities in actions were associated with transient increases in bilateral lateral occipital cortex. The overlap between these findings and the findings of the present study again points towards a role for visual and parahippocampal regions in event perception and binding event information, and furthermore suggests that insula may play a role in representing the temporal unfolding of events as well.

Thus, studies investigating brain activity during the passive perception of dynamic stimuli that unfold over time suggest that indeed, extrastriate visual areas, parahippocampal regions, the insula and the inferior frontal gyrus may play a role in perceiving events and their boundaries. However, these studies do not report activity in hippocampus. Differences between the present study and previous studies investigating event segmentation in the brain could be due to the (lack of) task employed by these studies, and by the stimuli that were used. Firstly, a task in which participants passively view stimuli without memorising them may not require hippocampal involvement (note that Swallow et al., 2011 does show an increase in hippocampal activity when retrieving information associated with event boundaries from memory and when retrieving information across event boundaries). Secondly, familiar naturalistic stimuli are segmented through the use of both bottom-up perceptual monitoring and top-down event models consisting of world knowledge. Hence, for events that are not novel, it may not be necessary to extract information about temporal regularities and transition probabilities between sub-events, as this information may already be consolidated into an event template. Hippocampus has been shown to play an important role in the initial extraction of this information, but its role in monitoring predictable events may be attenuated compared to its role in learning novel events.

5.7.4 Limitations and further research

As pointed out above, this was the first study to investigate retrospective duration estimation using a neuroimaging paradigm. Because of the fact that participants need to study and remember animations that will later on be reconstructed in terms of their duration, the number of items is limited, as there is a natural limit to how many items a participant can remember. This leads to an event related design that is by-definition relatively low in statistical power. This means that a higher number of participants would be necessary to obtain sufficient power. Due to practical limitations, however, this study was limited to 36 participants. Furthermore, because this is the first study to investigate retrospective duration estimation using fMRI, very few regions of interest could be defined a priori. Further studies could use the results of this study to guide focused investigations of task-related activity in predefined regions such as amygdala and lingual gyrus.

Secondly, design of the present study is a counter-balanced Latin square design, which means that each participant sees all conditions and all items, but not each condition of each item (one condition per item). As a result, comparisons on the individual level are not possible, because they are only meaningful on the counter-balanced group level. This also has implications for potential anatomical differences on the individual level: with the present design, individual differences in the anatomy of the MTL cannot be taken into account, because data are normalised to MNI space in order to perform analyses on the group level. To overcome this issue, participants would have to study each condition of each animation. However, this leads to confounds in the task, as participants may compare each condition of an item to each other, rather than providing an estimate based on their memory of the temporal unfolding of the animation, and representations of several conditions of one item could suffer from memory interference. Although these direct comparisons may also prove fruitful for investigating reasoning about the temporal unfolding of events, they are less informative about how we estimate the duration of a single, novel event (see Chapter 2 for a discussion about comparative duration judgements), and are less

comparable to the behavioural studies that have been reported in the previous chapters.

However, a benefit of employing a comparative design would be that it would allow for more detailed investigations of hippocampal representations. Using novel fMRI analysis methods such as multi voxel pattern analysis (MVPA) to investigate the representational similarity between the neuronal “signatures” of two stimuli in a certain area of the brain, such an approach may be a way forward in pinpointing the underlying organisational principles of hippocampal representations. For instance, is there a correlate of the number of segments or similarity between them in neuronal activity that can be observed across participants or across items? These questions cannot be answered with the current paradigm, but could be answered using a comparative design.

Furthermore, a second way of investigating the nature of hippocampal representations would be to make a direct comparison between encoding and retrieval of stimuli. Because of practical limitations (a trade-off between the number of participants that could be scanned for this study and the duration of the experiment), only the retrieval phase was performed in the scanner. However, performing both the learning task and the recognition memory and duration estimation task in the scanner would allow for comparisons between encoding and retrieval. Does encoding the stimuli in terms of their content elicit activity in the same brain regions as retrieving the information for duration estimation? How similar are the encoded representations to the retrieved representations? And do these patterns contain information about the number of segments and similarity? These questions and the proposed methods would provide the next step forward in investigating the role of hippocampus in encoding and retrieving information about the temporal unfolding of events.

Finally, the present study, like all other studies reported here up until this point, employed duration estimation on an arbitrary scale to obtain a dependent variable that reduces variability across participants (see Chapter 2 for a discussion). However, this means that providing a duration estimate does not only rely on retrieving event information, but also recruits cognitive mechanisms

concerned with scalar reasoning and decision making. Therefore, in order to directly tap into the mental representation of the temporal unfolding of events, a different task could be used. For instance, rather than asking for an arbitrary rating that might be confounded with decision strategies, one could also ask participants to simply mentally replay the animation. If the mental representation of the temporal unfolding of events would indeed be affected by event properties, then one would expect that the duration of the mental replay of the events would be similarly affected as the duration judgements shown in the previous studies. In terms of fMRI analysis methods, this paradigm would provide the advance of investigating correlations in brain activity between exposure, encoding and mental replay (see 6.6.2 for further discussion). In order to investigate whether it is possible to replicate the behavioural results shown here using a mental replay paradigm, allowing for generalisation across tasks, the next chapter presents a behavioural study employing mental replay as a dependent variable.

5.8 Conclusion

In sum, the present study is the first to show that regions in the MTL are involved in retrospective duration estimation. Furthermore, the results suggest that visual areas such as the lingual gyrus may play a role in representing visual event information that is relevant to duration estimation. These findings are in line with cognitive and anatomical models of MTL involvement in building associative representations. Moreover, the present results tentatively suggest that indeed, activity in these regions is modulated by event characteristics, similar to the modulation of duration estimates by event properties on a behavioural level. This direct relationship between the retrieval of event properties and the involvement of hippocampus in duration estimation points toward a more general role for hippocampus in learning regularities of stimuli, binding these in time and space, and retrieving these when retrieving temporal context. These observations may prove fruitful for further investigations of the nature of hippocampal encoding and

retrieval, as this is the first study to show a relationship between event content and the retrieval of temporal information.

Chapter 6

The effect of event structure on reconstructed duration: a mental reproduction study

6.1 Introduction

This chapter presents an experiment in which the effect of event structure on the mental reproduction of events was investigated. Similar to the experiments presented in the previous chapters, participants studied the content of animations over several exposures, but rather than providing a duration estimate on an arbitrary scale, they were asked to mentally replay the animation in exactly the way it occurred over its original time course, until its end point. This task more directly taps into the mental representation of the temporal unfolding of events that participants have. The data presented here shed light on the nature of event representations and the effect of event structure on duration representation, while also providing evidence for generalisation of these principles across tasks.

6.2 Mental replay of events

Throughout this thesis, a paradigm has been used in which participants were asked to provide a duration estimate on an arbitrary scale. Initially, this paradigm was used to decrease variability between subjects and to allow participants to judge duration without having to translate their 'sense' of duration into seconds, as they have no way to have this information available in a retrospective paradigm. However, as outlined in the Discussion of the previous chapter, this paradigm limits conclusions about the actual memory of the unfolding of the events. Decision-making processes rather than the memory representation of the underlying event structure could influence the choice of duration estimate. Therefore, the present study aimed to replicate and generalise the findings from Experiment 2, Chapter 2, using a paradigm in which participants were not asked to make scalar or magnitude decisions. Rather, they were asked to simply replay the

unfolding of the events in their mind, from memory, and press a button when they are done. This allowed us to assess whether indeed the memory representation of the unfolding of the events and the duration of this representation is affected by the underlying event structure, or whether it is rather the decision making process that is affected. Chapter 5 indeed showed involvement of MTL structures in the reconstruction of episodic memories during duration estimation, but the task was mediated by magnitude estimation on a scale, which can be susceptible to decision biases. Here we aimed to more directly investigate the duration of the mental reconstruction of the events.

Moreover, in Chapter 4, it was noted that reproduction measures in prospective paradigms tend to provide opposite results than numerical estimations of clock time. It was argued that interval reproduction tasks may be a measure of the speed of passage of time, thus, if many events took place in an interval, time was perceived as passing quickly (Liverence & Scholl, 2012), and thus more events lead to shorter reproductions than fewer events. These findings therefore raise the question of whether similar asymmetries may occur in retrospective paradigms. Perhaps, in mentally replaying the events from memory, more sub-events are re-played faster than fewer events, or perhaps, the rate of mental replay is entirely unrelated or independent from the original experience. This is an important question not only to understand the nature of the memory representations we extract from experience but also to illuminate its implications for applied cognition (e.g., witness testimony).

6.3 Experiment 8: Reconstructing the unfolding of events from memory

6.3.1 Research hypotheses and aims

Experiment 2 in Chapter 2 has shown that recognition latencies and duration estimates display a linear trend over conditions. Results from the regression analyses corroborated the independent effect of the number of event boundaries

and the similarity between sub-events on the duration ratings. The aim of the current experiment was to replicate these findings using a retrospective reproduction paradigm, allowing for generalisation of the results across tasks and for comparisons between different measures in retrospective paradigms. Firstly, it was predicted that if mental event reproduction match those underlying scalar judgements, the recognition latencies and mental reproductions of the temporal unfolding of events should show a main effect of condition and a positive linear trend over conditions. Secondly, it was predicted that as the number of perceived event boundaries increases, and as the perceived similarity decreases, reproduction times should increase.

6.4 Methods

6.4.1 Participants

Eighty-eight native English speakers, students at the University of York, participated for course credit, course requirement or a small monetary reward. Three participants were unable to perform the reproduction tasks (noncompliance). Three participants who showed poor memory performance were excluded in order to match recognition memory performance across conditions (<40% correct responses in any one of the conditions), and one additional participant was excluded to keep a balanced number of participants per list (excluding the participant with the next worst recognition memory score for that list; 44% correct in one condition). All participants displayed correct rejection of foils above 50% (i.e. false alarm rate <50%). As per design, the exclusion of participants with poor memory performance resulted in non-significant differences in correct recognition memory (basic: $M=91\%$, numerous: $M=90\%$, dissimilar: $M=90\%$; Friedman's test n.s.). Participants had normal or corrected-to-normal vision. This experiment was approved by the Ethics Committee of the Department of Psychology of the University of York. Participants provided informed consent and were debriefed after the study.

6.4.2 Materials

The 28 animation triads described in Experiment 2, Chapter 2 were also used in this experiment. The description of these materials can be found in section 2.72.

6.4.3 Design and procedure

The design of this study was similar to that of Experiment 2. The 28 stimulus triads were arranged into three different lists as was done for Experiment 2 (Latin square design). These lists were randomly assigned to participants. The experiment consisted of three tasks: study, recognition memory and mental reproduction. The study task was identical to that used in Experiment 2, except that participants were given the option to watch the animations again after the third cycle during learning. This was done to guarantee good memory encoding. Thus, after the third presentation cycle, participants were asked whether they were confident that they had learned the animations or whether they wanted to see the animations once more. If they responded positively, the program automatically cycled through the animations again in random order (21 participants chose this option and watched the animations a fourth time). If a participant was confident of having learned the animations, the stimulus presentation program moved on to the next task.

Participants then performed a recognition task, identical to the one presented in Experiment 2. From these data, the response latencies of YES-correct responses were analysed. Outliers above 3 SD from each participant's and each condition's mean response times were removed to obtain near normal distributions (less than 5% of the analysed data). Analyses conducted on either reaction times or log-transforms as dependent variables yielded the same pattern of results.

In the final task (unbeknown to the participants beforehand), participants were instructed to mentally reproduce (i.e., replay) the animation associated with the cue-frame in the way it exactly occurred over the original time course until it reached its original ending point (instructions cf. Boltz, 1995). Each trial in this

task started with the display of a cue-frame from those studied in the corresponding list. To minimise hand movements and time delays, participants were instructed to rest their index finger on the left-hand side button of a computer mouse throughout the task. Participants used their dominant hand to provide responses. Upon seeing the frame, they had to mentally replay the corresponding animation and click on the mouse as soon as they were done. Reproduction times were computed from the beginning of the cue frame presentation up to participants' mouse response. For each individual reproduction time, ratio scores were computed, as commonly done in psychophysical studies (cf. Boltz, 1995). These scores represent the ratio of reproduced duration to clock duration, which indexes the extent to which duration reproductions were larger or smaller than the actual duration (a ratio of 1 indicates no difference between reproduced and actual duration). Outliers above and below 3 SD from each participant's and each condition's mean scores were removed from the data (39 trials out of a total of 2268 cases). Note that analyses conducted on either raw ratio scores or log-transforms yielded the same pattern of results. For the regression analyses, we simply used the reproduction times in milliseconds as the dependent variable to conduct the same regression analyses as those in Experiment 2.

6.5 Results

6.5.1 Recognition memory

We expected that similar to Experiment 2, the response latencies in the recognition memory task would display a main effect of condition and a positive trend over conditions. The results of repeated measures ANOVAs with log-transformed response latencies aggregated up to the subject (F_1) or item (F_2) level as a dependent variable revealed significant main effects of condition ($F_1(2, 160)=3.04$, $p=.05$, $\eta_p^2=.04$; $F_2(2, 54)=5.11$, $p=.009$, $\eta_p^2=.16$) and significant linear trends ($F_1(1,80)=5.24$, $p=.03$, $\eta_p^2=.06$, $F_2(1, 27)=8.85$, $p=.006$, $\eta_p^2=.25$) (Figure 31). These

results replicate the findings of Experiment 2 and therefore suggest that encoded event properties are activated during recognition as a function of condition.

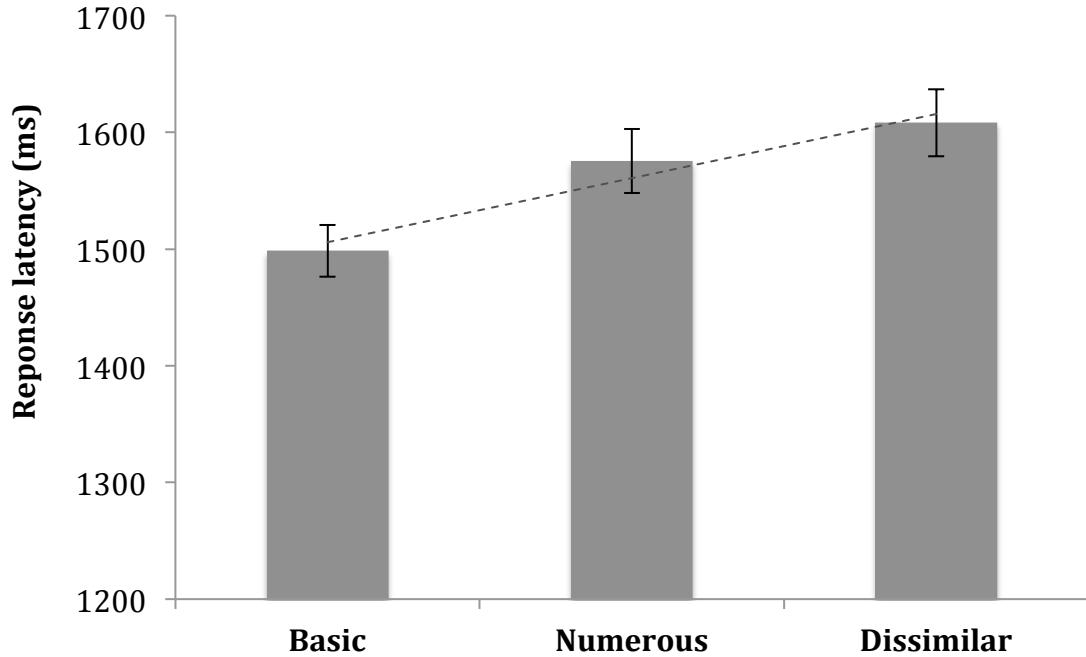


Figure 31. Response latencies over items. Error bars represent standard error. Trend line represents linear trend over conditions.

6.5.2 Mental replay

We expected that the time taken to mentally reproduce an event relative to its actual clock duration (ratio scores) would display a positive trend across conditions, due to more information being stored across conditions as a function of sub-event numbers and sub-event dissimilarity. Repeated measures ANOVAs with ratio scores (reproduced duration/actual duration) aggregated up to the subject (F_1) or item (F_2) level as a dependent variable revealed main effects of condition ($F_1(2, 160)=3.64, p=.03, \eta_p^2=.04$; $F_2(2, 54)= 4.16, p=.02, \eta_p^2=.13$) and significant linear trends ($F_1(1, 80)= 6.14, p=.02, \eta_p^2=.07$; $F_2(1, 27)= 6.28, p=.02, \eta_p^2=.19$) (Figure 32). These results suggest that event properties encoded in memory are

retrieved during mental reproduction, with reproduction times for events of the same clock duration increasing as a function of conditions' structural complexity.

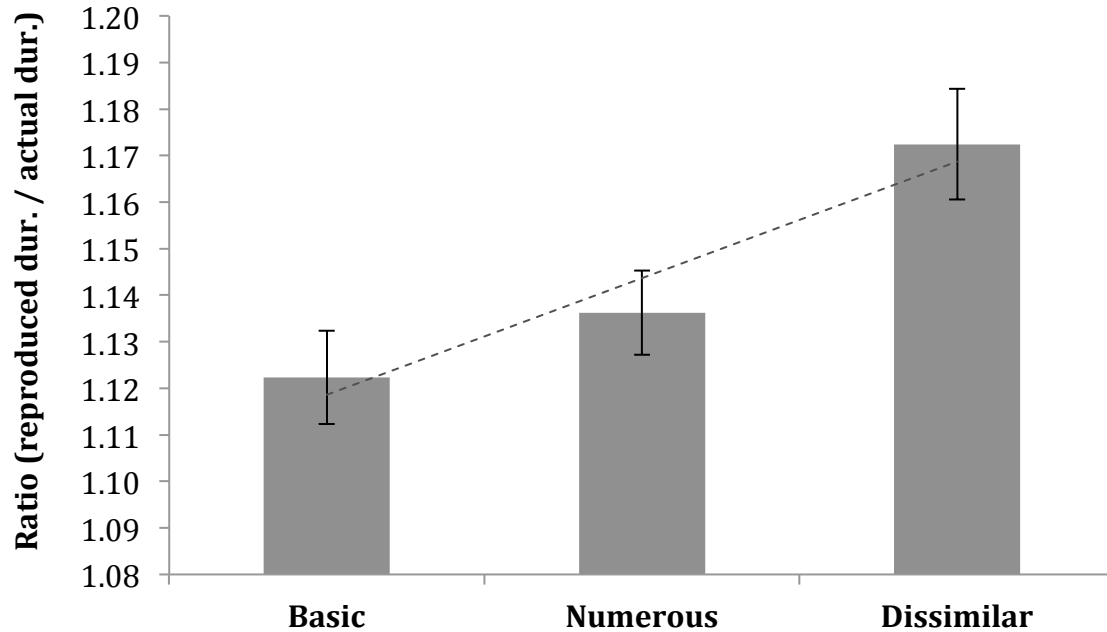


Figure 32. Proportional duration reproductions (reproduced duration/actual duration). Error bars represent standard errors after removing between-subject variability, see Loftus and Masson (1994). Trend line represents linear trend over conditions.

6.5.3 Regression results

To evaluate the contribution of sub-event and similarity properties in explaining reproduced durations, we conducted by-item hierarchical multiple regressions similar to those reported in Experiment 2. We examined the proportion of variance accounted for by sub-event and similarity scores (those obtained in the stimulus pre-tests), after clock duration was taken into account in the regression model. The first step of the regression model thus included clock duration as predictor and mean reproduction times per item as the dependent variable. We found that adding sub-event scores to this regression model significantly increased the proportion of variance accounted for: R increased from .81 to .87 ($F_{\text{change}}(1, 82) = 32.11, p < .0001$).

In the next step, we found that the addition of similarity scores to the model also significantly increased the proportion of variance accounted for: R increased from .87 to .88 ($F_{\text{change}}(1, 80) = 6.61, p = .01$). This pattern of significance remained, regardless of the order in which predictors were entered into the model. Table 3 provides the statistics for the full model. Thus, both the number of perceived sub-events and sub-event similarity play a role in duration reproductions once clock duration is taken into account, suggesting that participants mentally replayed the events based on the event structure encoded in memory.

Table 14. Regression coefficients for duration reproductions. Note: * indicates $p \leq .01$, and ** indicates $p < .001$

	Model	B	$SE\ B$	β
1.	Constant	3147.69	285.63	
	Clock duration	563.91	45.16	.81**
2.	Constant	2844.33	249.03	
	Clock duration	369.28	51.56	.53**
	Number of sub-events	317.47	56.03	.42**
3.	Constant	3654.06	396.45	
	Clock duration	355.72	50.14	.51**
	Number of sub-events	318.36	54.19	.42**
	Similarity	-156.76	60.97	-.14*

6.6 Discussion

6.6.1 Implications

The results presented here suggest that the encoded structure for events of the same clock duration, in particular sub-event and similarity structure, modulates our memory representation of events and their temporal unfolding: more sub-events and less similarity between them lead to longer recognition latencies and longer reproduced events. The reproduction task directly links encoded event structure to the mental replaying of the time course of the events, which, unlike comparative scalar estimates, is less susceptible to decision strategies. These results replicate and generalise the findings from the previous studies presented here, suggesting that sub-event and similarity structure mediate our memory representations of events and consequently, our reconstruction of duration.

Importantly, the present results also show that the findings from the previous studies reported here can be reproduced using a different task, suggesting that these findings are not task dependent and can be generalised to different tasks. One of the previously identified downsides of using a mental reproduction task is that these are known to be quite variable across participants (Grondin, 2008). Furthermore, for longer intervals reproduction tasks become more and more unreliable, and for very long durations they become tedious and unpractical (Grondin, 2008). However, for the durations of the present stimuli, this reproduction method has replicated the findings from Experiment 2, suggesting that this potentially more variable measure still picked up on the experimental manipulations.

Moreover, in prospective paradigms, interval reproduction tends to show contrasting results with other measures, such as interval comparison and numerical estimations. The present results are therefore important in highlighting that in retrospective paradigms, these asymmetries do not seem to occur. This is likely to be because the episodic memory representations underlying retrospective estimates are the same, as they were established during learning. In contrast, in

many paradigms, prospective estimates are dependent on working memory, without deeper encoding for later, and instead involve immediate recall. These working memory representations may lead to opposite results as those reported here, as higher numbers of events may make the speed of an interval seem faster when reproducing from working memory (see section 4.10.3).

Importantly, the task used here is not (primarily) temporal in nature. Given that participants were not asked to provide an interval reproduction, but rather they were asked to simply replay the animation in their mind from its beginning to its end, these results provide evidence for the effect of event structure on the representation of the unfolding of events, rather than on a purely temporal representation. In other words, it may indeed not be a purely *temporal* representation that is susceptible to the quantity and complexity of an event structure, but rather, as argued throughout this thesis, a multi-modal representation of the *content* of the event and its unfolding over time. This is in line with the findings from the previous chapters, including the findings from the neuroimaging study presented in Chapter 5 that suggest that indeed the information that is necessary to provide a duration estimate is obtained from regions encoding visual patterns that converge in hippocampus during duration reconstruction.

6.6.2 Further research

These findings are particularly important for further research investigating retrospective duration reconstruction using fMRI. As argued in the discussion of Chapter 5, it is currently unknown to what extent the encoded information about event boundaries during exposure overlaps with the information that is retrieved during duration estimation. Previous studies have identified that there is activity in certain areas of the brain during perception that is time-locked to the perceived event boundaries (Zacks, Braver, et al., 2001). To further investigate the nature of this information and to what extent this overlaps with retrieval, a paradigm in which the unfolding of events over time is retrieved may prove fruitful, as this

allows us to investigate whether the pattern of activity that is specific to the perception of an event boundary also occurs during the retrieval of that event boundary during the mental replay of the events.

Relatively new statistical methods for fMRI data analysis enable us to explore brain activity during dynamic event perception, and to further investigate the relationship between this activity and patterns of activity during retrieval. In particular, multivoxel pattern analysis (MVPA) is capable of taking into account the inter-relationships between the patterns of activity in all voxels measured at every time point within each subject (Spiers & Maguire, 2007). The pattern classifier that is trained on these inter-relationships is then applied to new fMRI data, which can tell whether or not the activity during encoding is predictive of activity during retrieval. Furthermore, MVPA can tell us more about what regions of the brain contain most voxels that show these discriminating patterns of activity, allowing us to infer what information is represented in these regions (Spiers & Maguire, 2007). Using these patterns of activity, one can identify overlap between the encoding and retrieval of event information. In particular, recent methods have been developed to estimate the degree of representational similarity between different conditions of a task (i.e. encoding and retrieval, or basic and dissimilar), and between different regions of the brain and between different subjects (Kriegeskorte, Mur, & Bandettini, 2008). Employing these methods is a way forward in investigating how the brain deals with dynamic, naturalistic stimuli by pinpointing what neural activity is induced by specific events within the continuous stream of ongoing activity (Spiers & Maguire, 2007). Using these analysis methods in combination with a mental replay paradigm may shed light on whether patterns of activity that are specific to the perception of an event boundary occur during both the perception and encoding of an event and during the retrieval or replay of an event, and may inform us about how information about events and event boundaries is represented by a network of regions.

6.7 Conclusion

In sum, the results presented here suggest that our memory representation of the temporal unfolding of events is affected by the number of sub-events and the similarity between them. These findings provide a generalisation of the effect observed in Chapters 2: more sub-events and less similarity between them lead to longer response latencies in recognition and to longer duration reproductions. These results corroborate the idea that it is indeed the memory representation of the unfolding of the events that is affected by event structure, rather than temporal reasoning or a decision process.

Chapter 7

Thesis summary, discussion and conclusions

7.1 Summary

In this thesis, a combination of behavioural and neuroimaging techniques was used to investigate the effect of event content on the memory representation of the temporal unfolding of events and duration reconstruction. In Chapter 2, retrospective duration estimation was investigated in Experiment 1 and 2 using a novel paradigm in which participants first studied animations of dynamic events over subsequent exposures. They then performed a recognition memory task, followed by a surprise duration judgement task. The results from these two studies indicate that the number of sub-events and the relative similarity between them both modulated duration estimates, over and above actual duration and over and above each other. Furthermore, these studies indicate that recognising a still frame belonging to an animation that has more sub-events and/or lower similarity between sub-events takes longer than recognising a still frame from a less 'complex' animation, suggesting that the content of the animation is indeed reflected in the memory representation and source memory of a paired-associate relationship between the memory representation of the events and the still frame.

In Chapter 3, the relationship between verbal encoding and duration estimation was explored in Experiment 3, and the same paradigm as above was used with verbal stimuli in Experiment 4 to investigate whether duration estimation based on verbal descriptions is affected by the same properties as estimating duration based on a memory representation of visual events. Experiment 3 indicates that surprisingly, the relationship between verbally encoded memories (operationalised as descriptions) and duration estimates is not as straightforward as assumed: although generally, there is a positive relationship between the number of words used to describe a memory and the duration estimate, it is not the case that this is captured in the number of verbs or the

number of different verbs. Taking the number of words out of the equation, Experiment 4 investigated duration estimation based on verbal descriptions of the animations used in Experiment 1. This study suggests that duration estimates based on verbal descriptions are affected by the number of sub-events, but not (or only to a very small extent) by similarity.

Together, the findings from Experiment 3 and 4 suggest two things: firstly, the lack of a clear relationship between verbal measures (e.g. the number of verbs, number of different verbs, etc.) and duration estimation, in combination with the strong correlation between the number of sub-events (as observed in the visual events, which formed the basis for the descriptions used here) suggests that people may go beyond the actual language used to describe the events when obtaining an event representation, thereby relying on the underlying event representation built during encoding. Secondly, the lack of an effect of similarity suggests that there might be systematic differences between obtaining an event representation and / or judging duration based on verbal descriptions of events and visual events. Further studies are necessary to investigate whether this is indeed the case, or whether this is an artefact of the stimuli used here.

In Chapter 4, two experiments were presented in which the effect of encoding content and monitoring time was investigated. In Experiment 5, participants were instructed to do the same task as in Experiment 2, but this time participants were told to also monitor the relative duration of the animations, as this would be required for a later duration task. In Experiment 6, participants performed the same task, but were instructed to monitor the relative duration of the animations and did not receive any instructions about remembering the content of the animations. The results from Experiment 5 showed that when participants attend to time and content, the number of event boundaries and similarity, much like Experiment 2, modulated duration estimates. However, Experiment 6 showed that when participants only pay attention to time, duration estimates were only correlated with the number of event boundaries and not by similarity. Together these findings suggest that the number of (coarse) segments may – to some extent – play a role during encoding of temporal unfolding, whereas

more fine-grained event boundaries and relative similarity between them only play a role when participants have a deeper encoding of event structure.

Chapter 5 presented Experiment 7, the first fMRI study to use a retrospective duration estimation paradigm. This study used exactly the same paradigm as Experiment 2 and was thus expected to replicate the behavioural findings of this experiment. Furthermore, the neuroimaging data was used to investigate the role of the medial temporal lobe (particularly left hippocampus) in retrospective duration estimation. The results suggest that activity in hippocampus showed the same modulation of conditions as the behavioural data. In addition, an exploration of the effect of each event characteristic revealed that indeed, areas in the MTL appeared to be affected by an increase in the number of sub-events and a decrease in the similarity between them. Furthermore, further whole brain explorations of the data revealed that hippocampus might play a role in binding multimodal information relevant to event properties, such as motion, colour and general patterns. The role of hippocampus in duration estimation puts forward the idea that hippocampus is involved in learning regularities of stimuli, binding these in time and space and retrieving these when reconstructing temporal unfolding.

Experiment 8 in Chapter 6 replicated and extended the previous findings using a different task. Rather than asking for a duration estimate, Experiment 8 aimed to directly tap into the mental representation of the temporal unfolding of events by asking participants to mentally replay the events and press a button when they are done. The findings showed that firstly, the effect of an increase in duration when more event boundaries with less similarity between them need to be reconstructed could also be observed with a different, non-temporal task, suggesting that the findings presented here are reproducible and generalizable. Secondly, the findings suggested that indeed, the actual mental representation of the temporal unfolding of events is affected by event structure, rather than (just) temporal reasoning or decision-making.

7.2 Aims and implications

7.2.1 General aims

The aim of this research was to investigate the effect of event content on the memory representation of the temporal unfolding of events and duration reconstruction. In particular, two aspects of event content were investigated: firstly, the number of event boundaries or sub-events as identified by observers, and secondly, the perceived relative similarity between these sub-events. This approach aimed to bring together two lines of research that have not previously been united in order to investigate how event content drives duration reconstruction, namely the field of research into time perception and research investigating event perception and memory.

Despite the intuitive connection between encoding event content and reconstructing events and their duration from memory, this has not been investigated before with systematic rigour. In particular, retrospective duration estimation has been an unpopular paradigm because of its practical constraints: as soon as participants are aware of the experimental aim (i.e. the investigation of duration representation), they are likely to actively monitor time which alters the retrieval process from reconstructing the event to retrieving the temporal measure. Furthermore, no studies have previously investigated these issues using neuroimaging methods. In particular the restricted number of items can be problematic: as participants can only remember a certain number of stimuli, studies using a reconstruction paradigm are easily underpowered.

Furthermore, although some studies have previously pointed towards the number of “changes” (cf. Fraisse, 1963; Poynter, 1989) or the number of “stimuli” (cf. Ornstein, 1969) as an aspect of event content that affects duration estimation, the present studies have taken recent insights from the event perception literature to further pinpoint the characteristics of dynamic events that modulate duration estimates. The stimuli used throughout this thesis have been developed employing insights into what event properties drive the perception of event boundaries,

leading to three conditions. Moreover, rather than solely interrogating the data using this potentially arbitrary division into conditions, the present studies have also employed naive, perceptual ratings of the number of event boundaries and relative similarity between segments to investigate the relative contribution of these characteristics on duration estimates and event reproduction.

7.2.2 Implications for theories of time perception

In Chapter 1, several cognitive models of duration perception were introduced. These models all assume that there is some relationship between stimulus characteristics and duration estimation in retrospective paradigms, but all differ in what they propose exactly. Change based models propose that it is the number of perceived changes that defines the duration estimate, whereas Storage Size models and Processing Effort models assume that the number of stimuli and ease with which they could be encoded determines the attributed duration. However, none of these models have investigated which properties of stimuli determine the amount of encoded information or the efficiency with which this information can be encoded. The present results suggest that models that take the number of changes as their index of the amount of information are overly simple: over and above the number of changes, the similarity between them affects retrospective duration estimates. The present findings are consistent with Contextual Change models, which propose that duration estimation is guided by the variety in processing, and moreover extend these by being the first to pinpoint two perceptual characteristics of events that explain the observed effects of “Contextual Change”.

Importantly, the use of naturalistic event perception and segmentation is new to the field of time perception research: manipulating natural event segmentation and using a bottom-up approach of asking naïve participants to provide counts of the number of sub-events and ratings of the relative similarity between them is novel, and proves to be an important step forward in investigating the perceptual characteristics that drive effects of subjective duration in (retrospective) duration estimation. As outlined in Chapter 1, the term ‘event’ was

previously used in time perception research but was used to denote any kind of stimulus presentation, and the word 'segmentation' was used to describe any task-specific grouping mechanisms. For instance, previous studies investigated the number of salient words in a list of words (Poynter, 1983; Zakay et al., 1994) or the number or type of tasks instructed to participants (Block & Reed, 1978), essentially investigating the effect of pre-determined stimulus chunks on time estimation.

Although these studies are related to 'segmentation' and 'event' perception in the sense of providing mechanisms for encoding the stimuli that may or may not aid participants, they bear little relation to perceiving and encoding a dynamic sequence of events in an unguided manner, on the basis of the internal causal structure of events. The present studies pre-tested a separate set of participants to segment the stimuli in causally bound event units. The critical point is that this dynamic event unitisation predicted what naïve participants did during recall. Thus, the participants are encoding and retrieving natural event segments guided by the causal relations of the natural world, not arbitrary chunks imposed by the experiment paradigm. Therefore, while it might not be entirely surprising that chunking or segmentation aids memory and time estimation, the critical point is that people inadvertently structure and encode dynamic sequences of events according to cause-effect relations and the perceived similarity structure in the world.

The novelty of the present studies is thus that they inform us about the content of event memories and their relationship to temporal memory. The mental replay study (Experiment 8), which involves no duration estimation decision at all, directly taps into the representation of the temporal unfolding of the events. This is a fundamental issue that time estimation studies have not investigated before: as outlined in Chapter 1, many of the time estimation studies have been occupied with investigating these decision making processes rather than investigating the content and nature of memory representations themselves. Thus, the studies presented here provide evidence for the role of memory for event content in retrospective duration estimation, and that this memory representation is driven by event characteristics such as the number of sub-events and the similarity between them.

The results of Chapter 4 also contribute to the understanding of prospective paradigms, which have traditionally been explained mostly in terms of attention (with the exception of a few recent studies, e.g., Waldum & Sahakyan, 2013). The results from Experiment 6 suggested that even when participants pay attention to time only and are exposed to the stimulus several times, duration ratings in a prospective paradigm without explicit memory for content appear to also be biased by aspects of event segmentation, which appear to contribute to the time keeping process. This suggests that prospective timing not only depends on attention to time, but also on content. Similarly, attention theories take dual task paradigms to involve more cognitive load and therefore, predict detrimental effects of shared resources on time keeping. However, a decrease in duration attribution as an effect of dual tasking can be precluded when this attention is directed to content, as more encoded content leads to longer duration estimates. The prospective dual-task paradigm (Experiment 5) in fact did not show a detriment in attributed duration relative to a purely prospective task. Instead, on average, duration ratings were similar, but increased as an effect of condition for the dual task paradigm. These findings suggest that attending to content leads to more encoded information about the event when both time and content are attended to, precluding a detriment in attributed duration. These findings argue against the traditional view that prospective and retrospective duration estimates are differentially affected by attention and memory respectively (e.g., Block et al., 2010). Moreover, they suggest that both prospective and retrospective duration estimates are affected by event structure. Further comparisons between prospective and retrospective paradigms showed that attending to time does contribute to the amount of accumulated time, with prospective paradigms leading to longer duration estimates. These findings are in line with previous studies investigating the effect of attention to time on duration estimates (e.g., Brown & Boltz, 2002).

7.2.3 Implications for the statistical learning literature

One of the aims of these studies was to investigate how we reconstruct the duration of novel events. For this reason, a learning paradigm was employed in which animations of novel events were used that resemble naturally occurring events in the real world in that they are based on the causal and contingency relations found in the real world, but that are novel enough to avoid top-down event processing (as this would affect for instance the perception of human actions). Perceiving and encoding such contingencies is critical when we first encounter and learn about naturally occurring events as children, as suggested by statistical learning studies. The binding of event properties into memory schemas depends on recurrent patterns of co-variation between these properties in our experience. These patterns are known to lead to schema extraction in cognitive development (Sloutsky, 2003; Fiser & Aslin, 2005; Orbán et al., 2008). Therefore, the present thesis sheds light on what aspects of causal events we encode with repeated exposure, and how these aspects modulate the memory representation and replaying of the temporal unfolding of events. These insights are important for learning about the world and building representations of the unfolding of the events that we have encountered (schema formation in semantic memory).

The results contribute to this line of research by suggesting that for spontaneously encoded event structures of causal events that we have encountered equally often, participants provide duration estimates based on the encoded segmentation and similarity structure of the events. Moreover, the results suggest that the nature of the encoded segments (e.g. colour, shape, motion) and their similarity to the contexts in which they occur play an important role not just in duration estimation but also in recognition memory and in the mental replay of the unfolding of an event. These findings are important, because contextual similarity (or dissimilarity) has long been recognised to modulate category formation during development (Sloutsky, 2003) and stimulus segmentation in statistical learning (Avrahami & Kareev, 1994; Gómez & Gerken, 2000). These studies suggest that similarity is fundamental in determining segment perception

and learning in general (Goldstone, 1994). This previous research together with the present results therefore argues for common structuring principles shaping memory formation and the resulting memory content.

7.2.4 Implications for the event memory literature

Previous studies in the domain of event perception and event memory have already suggested that the way in which we segment the stream of on-going activity into events is what forms the basis for our memory of what happened (Zacks et al., 2007). For instance, more fine-grained event segmentations lead to more stored information (Hanson & Hirst, 1989), and information associated with event boundaries is more likely to be encoded in episodic memory than information that is part of the event but not associated with a boundary (Swallow et al., 2011). The present findings contribute to this research by showing that—as pointed out above—event segmentation and memory for event structure affect duration estimates and mental replay of events. This direct relationship between the structure of events and representations of duration and temporal unfolding has not previously been investigated in this field (although language research has pointed in this direction; see below). Moreover, these novel findings stress the importance of event structure for more general cognitive processes. By using duration estimation and the duration of mental replay as a proxy for tapping into mental representations of event unfolding, the studies presented here illustrate that event structure is pivotal to the formation of event representations.

This is to some extent further supported by the findings from the recognition memory task, which in each of the experiments show a pattern similar to the duration ratings, suggesting that there may indeed be a relationship between how the event are encoded and the duration estimates. However, these results need to be interpreted with care. These response latencies cannot be explained in terms of the number of segments and similarity (i.e. regression analyses with the event properties did not yield significant proportions of variance explained) suggesting that the relationship between the latencies and the event properties is

not as evident as for the duration ratings. However the global sensitivity to the conditions suggests that event content may still to some extent affect recognition latencies. It may however not be necessary to reactivate all event information, as not all information may be relevant to the decision making process. A shallower level of retrieval may be sufficient that reflects the global structural aspects of the animations, but not detailed enough to reflect all individual event properties.

Furthermore, it may not be necessary to retrieve or mentally replay the animation as a whole in order to make a recognition decision. Moreover, as participants were instructed to respond as quickly and accurately as they could, they were given less time to consider the animations in detail. Also, although the participants were not made aware of this fact, the cue-frames that elicited YES-responses were taken from the beginning of the animation and had been seen during the study phase. Hence, it may not be necessary to reactivate the whole animation with all of its event boundaries in detail to make the recognition decision, while it is necessary to do so in order to make a duration estimate. Furthermore, reusing the cue-frames in the recognition tasks suggests that it may have even been possible to make these recognition decisions purely based on the familiarity of the cue-frames. However, because the foils required participants to retrieve more details of the unfolding of the events, participants were encouraged and expected to activate at least some event information. It may however be the case that for correctly recognising a cue frame, reactivating only a gist-like representation of the event is sufficient, which may explain why the general conditions are reflected in the response latencies but not the individual item-by-item event properties.

In the current studies, only NO-responses were obtained for later points in the animation. This limits any conclusions about whether the speed of recognising an event situation at a later point does or does not reflect individual event properties, as NO-responses are known to be highly variable and it is unclear what underlying cognitive processes they reflect. Further studies could investigate whether YES-cue frames from later on in an animation show a stronger relationship with the individual event properties.

Thus, the present studies shed some light on the relationship between event content and recognition memory. Although not demonstrated exhaustively, they suggest that cue-frames from more complex events as defined by the conditions take longer to be recognised. This is in line with findings from previous research. As shown by Swallow and colleagues (Swallow et al., 2011), information that is associated with an event boundary is more likely to be stored in episodic memory. Encountering more event boundaries in an animation may thus lead to more encoded information about the events and thus, more associated information. Previous studies in the domain of language have shown that indeed verb phrases that bring to mind more and more diverse associated information are recognised slower than phrases that afford less associated information (Coll-Florit & Gennari, 2011). This line of research may thus be a promising avenue to further address to what extent cognitive processes are affected by event structure, and to what extent episodic event memory resembles and differs from the information stored in semantic schemas that are employed in language comprehension.

7.2.5 Implications for the semantic memory literature

As argued above, the role of similarity structure in estimating duration has been hypothesised based on previous findings from language. Both results from studies using verb phrases and the results presented here from visual events suggest that similarity and sub-event structure play a role in event and duration reconstruction. However, this leads to the question whether the conception of similarity in language and visual encoding is the same. Semantic memory contains representations that have been built over many exposures in many different contexts, leading to an abstract representation that is independent of the actual event experience. This semantic representation is likely to consist of a definition of the event in terms of its features, for instance its typical agents and patients and its typical location, as determined by co-occurrences (Ferretti, McRae, & Hatherell, 2001; Hare, Jones, Thomson, Kelly, & McRae, 2009). This means that contextual

diversity leads to more associated information, as co-occurrences may be more distributed for events that occur in more diverse contexts.

This contextual diversity is less apparent in the learning paradigm used in the retrospective experiments presented here. In these visual studies, similarity structure is mainly operationalised as the dissimilarity of subsequent changes in a series of (sub-)events within an animation. In particular, the only contextual information that participants can obtain about the distribution of events is based on their occurrence with regard to their preceding and subsequent events within the animation. This gives rise to some contextual diversity, albeit restricted to a few occurrences and only within the same animation. Thus, limited information can be obtained about the occurrence of a certain event in general, as participants do not see the same events occur in different contexts (e.g., a similar animation with the same distribution of events but in a different order, or the same distribution of events in a different spatial context). Although the present results point to similarities between event perception and episodic representations on the one hand and semantic memory and schema formation on the other hand, further research is necessary to explore to what extent both rely on extraction of the same features, whether the conception of these features is the same across vision and language and whether these share underlying cognitive and neural mechanisms.

In relation to the findings from the verb phrase studies cited above (Coll-Florit & Gennari, 2011), the results from Experiment 4 were thus surprising in that they did not reveal an effect of similarity when estimating duration based on a memory representation of the unfolding of events as obtained from a description. However, as pointed out above, there are some methodological concerns with the study that need to be taken into account. For instance, Experiment 3 showed that there was a strong correlation between the number of words used naturally to describe an animation and the duration estimates. Because the number of words was kept constant between conditions in Experiment 4, it may be the case that the descriptions – although they were accurate and complete – did not communicate the event structure clearly. Thus, further studies are necessary to ascertain whether indeed, similarity plays no or only a minor role in reconstructing duration

from a verbal description.

7.2.6 Implications for the hippocampus literature

The results from the neuroimaging study presented in Chapter 5 have implicated hippocampus and surrounding areas in the MTL in retrospective duration estimation. This finding is novel, as studies have not investigated retrospective duration estimation before using a neuroimaging paradigm and as such contributes to previous research by shedding light on the areas involved in and the nature of the memory representations underlying this process. These findings emphasise the constructive nature of memory (cf. Buckner & Wheeler, 2001; Slotnick & Schacter, 2006; Addis et al., 2007), suggesting that relevant elements of the event structure are reactivated and reintegrated during this process. Furthermore, the present data suggest that activity in hippocampus is modulated by the number of sub-events and similarity between them (as conveyed by our conditions). This relationship between retrospective duration estimation, event structure and activity in the MTL has not been shown before, and as such provides the first evidence for the role of the content of an episodic memory and the amount of processing necessary to reconstruct it during duration estimation.

However, as pointed out above, a limitation of this study is that although it implicated hippocampus in retrospective duration estimation and although it showed that activity in hippocampus is modulated by event structure, it does not inform us about the exact nature of the underlying computations that hippocampus is involved in during duration estimation. Based on previous research, it is thought that hippocampus plays a role in recollecting event information, thinking about events in the past and projecting into the future, encoding and reasoning about temporal order and encoding temporal anchoring. All of these processes (or their retrieval counterparts) are likely to play a role in retrospective duration estimation, and the data presented here cannot distinguish between these processes. Furthermore, given that fMRI measures the average BOLD response in a voxel, it is not possible to distinguish between the contributions of neurons with

different sensitivities using the current fMRI paradigm, making it hard to directly link results from an fMRI study to neurophysiological findings, such as those from studies investigating time cells.

Moreover, the present results are in line with theories that propose that hippocampus has distinct properties that allow it to serve as a convergence site for event information (Howard Eichenbaum, 2004; Paz et al., 2010). Because hippocampal neurons are capable of reflecting spatial and temporal aspects of events, they are capable of providing the spatiotemporal context of events. This property of hippocampus may prove critical to its role in retrospective duration estimation. To further corroborate this, further research is necessary. In particular, comparing brain activity between the study phase (i.e., learning novel events) and retrieval phase during a retrospective duration estimation task may inform us about the overlap between the (nature of the) encoded event information and the information retrieved during duration estimation. Furthermore, comparing prospective and retrospective duration estimation may shed light on whether and to what extent encoded temporal information as measured with a prospective paradigm overlaps with the retrieved information during a retrospective paradigm.

7.3 Limitations and further research

As argued above, the mechanisms proposed here may be applicable to learning novel events, for instance to the way that children learn about new events by monitoring perceptual cues such as changes and monitoring co-occurrences of situations. This does not mean that these mechanisms are also applicable when we encounter events that we are familiar with. As experienced perceivers, we are likely to in many cases have some top-down knowledge and predictions about event structures. The present study is limited in exploring the interplay between bottom-up perceptual features and top-down knowledge: it does not take into account situations in which knowledge that we have about other events in the world affects our event perception (for instance, when abstract shapes interacting with each other could be interpreted or described as humans performing actions).

Thus, an open question here is to what extent the behavioural findings presented here translate to naturalistic situations, particularly situations in which we encounter an event that is similar to other events that we have already experienced. However, as argued above, these insights are nevertheless valuable for statistical learning, as they shed light onto what kind of information is spontaneously extracted and retained from a stream of dynamic events.

Given that statistical learning relies on the brain's sensitivity to co-occurrence probabilities, another open question is to what extent these mechanisms apply to a single experience of a novel event. In the studies presented here, participants study the events over several exposures, building an abstract representation of the event content. The identification of segments can then be 'fine tuned' over these exposures by identifying co-occurring elements and by identifying the relevant grain for the current purpose (compare this to for example listening to speech in an unknown language: the stream of sounds does not appear to have any boundaries until one is able to group together the co-occurring sounds into words). However, when we only encounter an event once (e.g., for the very first time) it is unclear whether and how we identify segments and how similarity becomes apparent. It is likely that this process heavily relies on bottom-up processing of perceptual features, and may thus be largely driven by for instance salient changes (e.g., an odd-ball effect). Although it is outside of the scope of the present thesis, there is some evidence that these salient changes lead to temporal illusions, such as subjective lengthening of time when an odd-ball stimulus is presented in a train of items (Pariyadath & Eagleman, 2007). It would thus be interesting for both research into statistical learning and event perception as well as time perception research to pursue this line of research, investigating what properties of novel dynamic events drive these temporal illusions at first exposure.

Another open question is to what extent these findings are susceptible to individual differences. Individuals may differ in their experience with certain events. For instance, studies have suggested that the experience of the passage of time depends on an individual's familiarity with performing a certain task: routine tasks appear to go quicker than non-routine tasks (Avni-Babad & Ritov, 2003). This

may be due to people's grain of segmentation: a novel task may require remembering each individual sub-event, whereas for routine tasks these may be integrated into a larger event based on the high co-occurrence of the sub-events. Hence, duration estimates and the duration of mental replay may depend on an individual's familiarity with certain events.

This may also be true for cross-cultural differences. In particular, the way events are described by language can substantially differ between languages. For instance, while English only has one verb for placing an item onto a surface, namely 'to put', Dutch distinguishes between 'to put in a standing position' and 'to put in a laying down position' based on the object's position as it is placed on the surface (Flecken & Van Bergen, 2014). For an English speaker, a stream of events containing both kinds of 'putting'-events may be remembered as very similar 'putting'-events that are repeated (i.e., comparable to the numerous condition presented here), whereas for a Dutch speaker, this may be remembered as a sequence of 'putting to stand' and 'putting to lay' events (i.e., comparable to the dissimilar condition). This question whether differences in how language describes the world around us affects our perception (not only of events, but also for instance colour, e.g., Winawer et al., 2007) has been around for centuries and is still at the centre of debates about the relationship between language and perception. Extending the studies presented here to cross-cultural situations, employing for instance these cross-linguistic differences in describing 'putting'-events, may shed light on whether event perception, as observed by differences in reconstructed duration and mental replay, differs between speakers of different languages.

There are other factors that are not taken into account or not manipulated in these studies that may affect time perception and perhaps duration reconstruction on an individual level. In particular, individual differences may become apparent in prospective and dual tasks. For instance, an individual's working memory and attentional control have been shown to affect the experience of time passing: time appears to go faster for people with a higher working-memory capacity and more attentional control, as more resources can be devoted to time keeping (Woehrle & Magliano, 2012; Zakay & Block, 2004). This also

suggests that a decline in working memory, for instance as an effect of aging, may affect the experience of time. Furthermore, individuals may generally differ in their attitude towards time, with some people being more focused on time and time keeping than others (Zimbardo & Boyd, 1999). Finally, a pilot study using a prospective paradigm with the animations used throughout this thesis has shown that emotion may affect time perception: people who are primed to be in a happy mood may perceive stimuli as longer than people who are primed to be sad (Bono, 2014, unpublished MSc thesis). Although these individual differences should not bias the results presented throughout this thesis (as there are all averaged on a group level), it might nevertheless be interesting for further research to explore these differences, in particular to explore whether any of them affect both time keeping and duration reconstruction, shedding light on any interplay between the two.

7.4 Conclusions

Together, the findings from the eight studies presented in this thesis suggest that event structure plays a role in duration representation, both when measured through duration estimation and through mental replay of events. The behavioural results from Experiments 1, 2, 5, 7 and 8 all show that both the number of perceived sub-events and the similarity between them affect duration estimates and mental reproductions: higher numbers of sub-events and less similarity between them lead to higher estimates and longer reproductions. These findings are accompanied by consistent evidence from recognition memory tasks, suggesting that there is indeed a relationship between these event characteristics and the memory representation of the events, as shown by the increase in recognition latencies for animations with more sub-events and / or less similarity between them. Experiment 6 shows that these findings indeed appear to rely on the encoded event structure, as the role of event structure is diminished when participants do not attend to event content at all but only to duration. However, interestingly, on a coarse level, the number of event boundaries may guide the

encoding of duration and events more generally, even when participants do not attend to content. The findings from Experiment 3 and 4 contribute to the debate by showing that obtaining an event representation from language and estimating its duration may rely on both similar and different features as doing so based on a memory representation of visual events, opening up a new potential avenue of research. The neuroimaging findings from Experiment 7 presented in this thesis are the first to show involvement of the hippocampus in retrospective duration estimation, and to show that this involvement is modulated by the event properties under investigation here.

Appendix A.

Stimuli used in Experiment 4. Order of conditions: basic, numerous, dissimilar.

- Item 1 The red ball comes in from the bottom right hand corner and bumps its way up from box to box diagonally before leaving from the top. The red circle travels upwards, hitting blocks diagonally. Once it hits the blocks it knocks the blocks off the grid. The red ball bounces diagonally and knocks into every other square. The ball knocks three squares out and two squares disappeared.
- Item 2 The ball starts by moving in from the top left hand corner and rolls down on each of the three slopes falling off the bottom of the last slope and off the screen. The circle starts at the top of the first ramp and rolls down like a ball. When a ramp is used it slides off the screen. The ball falls off the screen. A ball rolls down the first line and then the line disappears. After the ball leaves the 2nd line, the line turns orange. When it leaves last line, the line disappears.
- Item 3 The far right circle swings and hits the middle circle, which hits the far left circle. The reverse then happens for a total of six hits. The final two hits are significantly shorter in distance. Three circle-shaped chimes bounce from side-to-side against each other. Each time the middle chime is struck by the outer chimes, it changes colour from black to red and causes the other chime to move. The circles swing like a Newton's cradle. As they hit, the middle circle turns red, then the left circle turns green, then the right circle turns blue, the middle turns yellow and then black.
- Item 4 Five circles drop straight down from out of the top of the frame, one at a time and land in the squares in the 1, 4, 3, 2, 5 order from left to right and then stop. Five balls in total. First ball drops from the top of the screen, enters the first box and turns green. Balls fall into box 4, 3, 2 and 5 after this, also turning green. Circles drop into each box in the following order: one, four, three, two, five. The circles turn green in the top three boxes. The boxes turn blue in the bottom two boxes.
- Item 5 The box rolls from left to right, it falls into the holes, and rolls back out of them both times, stopping at the far right side of the screen. The box moves to the right. When it falls into the two depressions, it knocks the bottom line off the screen and continues its path across the screen. The square rolls from left to right. When it drops into one depression, the bottom opens and flaps. When it drops into the other depression,

- the bottom falls out.
- Item 6 The circle hits four squares, starting with the top one, and moving in a clockwise direction. The circle returns to the centre of the screen after hitting each square.
 With a centre reference the circle bounces off the squares one at a time, starting with the top and moving clockwise. The squares respond by bouncing out and then back.
 The circle hits each square in a clockwise manner starting at the top. On impact, the first square moves up and down. The next spins, the next moves up and down, and the last spins.
- Item 7 The black box opens up and the circles roll out, they roll around and roll back inside, and then the box closes again.
 The circles float out of the opening of the box and float around, changing colour from black, blue, red, yellow then finally green when they go back into the box.
 The black circles leave the box, become black squares, then blue squares, then blue lines, and after returning to the main box, become red lines.
- Item 8 The first circle moves from left to right passing over the other circles, causing them to disappear as they are passed over. The circle then moves off the screen to the right.
 The first ball moves to the right, covering the second ball. It then grows bigger, and continues moving to the right. Each time it covers a ball, it grows in size.
 Each circle from the left sequentially moves to absorb the circle on the right and then increases in size or changes colour from black to purple to orange as it moves across the screen.
- Item 9 The ball rolls smoothly down the staircase, hitting each step along the way, until it reaches the bottom of the steps.
 The ball rolls down the staircase from step to step. Each horizontal line falls straight down after being rolled on by the ball.
 The ball rolls down the steps knocking the first ledge down and turning the next ledge red. It alternates between these two effects.
- Item 10 Several different squares descend down on the screen. The squares land on the bottom of the screen and form into five different vertical columns of varying height.
 Boxes fall from the top of the screen forming stacks, changing colour when they hit the bottom; red, orange, turquoise, blue and purple from left to right, turning black again afterwards.
 Squares rain down into irregular stacks, hitting bottom with a change of either colour or shape, becoming either a circle or a coloured square before reverting back into black squares.
- Item 11 The circle moves upwards and off the screen, then reappears at the bottom of the diagonal column to the right, it then continues until it travels each diagonal column.

- The circle starts at the left top track and moves up the track. It warps to the next lane and turns into a square. This pattern continues, alternating between circle and square.
- A shape travels up each diagonal path starting at the bottom each time. It changes shape and colour in a pattern, black circle, black square, red square, red circle, repeat.
- Item 12 The square on the far left hits the second square, causing that square to push the third square, which hits the last square, which stops at the edge.
- The first box hits the second and turns blue; the second hits the third turning blue. The pattern continues till the last square touches the edge.
- The left square twirls and knocks into the second square, which turns blue. This hits the third square turning it blue. The third square hits fourth, which stays black.
- Item 13 The circle spirals around a centre point in a clockwise manner always staying in the same orbit.
- The circle moves in 6 small circles, changing colour on each rotation, through black, purple, blue and black again.
- The circle alternates between rotating clockwise and counter clockwise while changing colours: black, burgundy, black, burgundy, blue.
- Item 14 The triangle transforms into a square. Then reverses back into the triangle. Then the triangle transforms into a square and back to the triangle.
- The black triangle turns into a blue square, which turns into a green pentagon, which then turns into a larger red triangle, then back into the smaller black triangle.
- The black triangle morphs into a blue square, then into a green triangle. It then becomes a red square, then returns to the original shape: a black triangle.
- Item 15 The red box moves across the black grid, starting in the upper left corner. It moves horizontally going across each row, until it reaches the final box of the grid.
- Two red squares move separately from top left grid along the perimeter until they meet again at the bottom right grid, separate again and return to the top left grid.
- The border boxes flash red in a clockwise and anti-clockwise direction until they meet at the top left corner. The columns of boxes turn red and then blue and disappear.

References

- Adams, R. D. (1977). Intervening stimulus effects on category judgments of duration. *Perception & Psychophysics*, 21(6), 527–534.
<http://doi.org/10.3758/BF03198733>
- Addis, D. R., & Schacter, D. L. (2008). Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus*, 18(2), 227–237.
<http://doi.org/10.1002/hipo.20405>
- Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7), 1363–1377.
- Allan, L. G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 31(1), 24–31.
- Allan, L. G. (1978). Comments on current ratio-setting models for time perception. *Perception & Psychophysics*, 24(5), 444–450.
- Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, 26(5), 340–354. <http://doi.org/10.1093/acprof:oso/9780199228768.003.0021>
- Amaral, D. G. (1999). Introduction: What is where in the medial temporal lobe? *Hippocampus*, 9(1), 1–6. [http://doi.org/10.1002/\(SICI\)1098-1063\(1999\)9:1<1::AID-HIPO1>3.0.CO;2-T](http://doi.org/10.1002/(SICI)1098-1063(1999)9:1<1::AID-HIPO1>3.0.CO;2-T)
- Avni-Babad, D., & Ritov, I. (2003). Routine and the perception of time. *Journal of Experimental Psychology: General*, 132(4), 543–550.
<http://doi.org/10.1037/0096-3445.132.4.543>
- Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition*, 53(3), 239–261. [http://doi.org/10.1016/0010-0277\(94\)90050-7](http://doi.org/10.1016/0010-0277(94)90050-7)
- Bellezza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 990–997. <http://doi.org/10.1037/0278-7393.15.5.990>
- Berg, M. (1979). Temporal duration as a function of information processing. *Perceptual and Motor Skills*, 49(3), 988–990.
<http://doi.org/10.2466/pms.1979.49.3.988>

- Block, R. A. (1974). Memory and the experience of duration in retrospect. *Memory & Cognition*, 2(1), 153–60. <http://doi.org/10.3758/BF03197508>
- Block, R. A. (1990). Models of psychological time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 1–35). Lawrence Erlbaum Associates.
- Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap*. (pp. 141–152). Dordrecht, Netherlands: Kluwer Academic.
- 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. <http://doi.org/10.1016/j.actpsy.2010.03.006>
- Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning & Memory*, 4(6), 656–665. <http://doi.org/10.1037/0278-7393.4.6.656>
- Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, 4(2), 184–197. <http://doi.org/10.3758/BF03209393>
- Block, R. A., & Zakay, D. (2001). Psychological time at the millennium: Some past, present, future, and interdisciplinary issues. In *Time: Perspectives at the millennium (The study of time X)* (pp. 157–173).
- Block, R. A., & Zakay, D. (2008). Timing and remembering the past, the present, and the future. In S. Grondin (Ed.), *Psychology of time* (pp. 367–394). Emerald.
- Boltz, M. G. (1992). The remembering of auditory event durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 938–956. <http://doi.org/10.1037//0278-7393.18.5.938>
- Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics*, 57(7), 1080–1096. <http://doi.org/10.3758/BF03205466>
- Boltz, M. G. (2005). Duration judgments of naturalistic events in the auditory and visual modalities. *Perception & Psychophysics*, 67(8), 1362–1375. <http://doi.org/10.3758/BF03193641>
- Bono, S. (2014). *Unpublished MSc Thesis*. University of York.

- Botzung, A., Denkova, E., & Manning, L. (2008). Experiencing past and future personal events: Functional neuroimaging evidence on the neural bases of mental time travel. *Brain and Cognition*, 66(2), 202–212.
<http://doi.org/10.1016/j.bandc.2007.07.011>
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology. General*, 138(4), 487–502.
<http://doi.org/10.1167/8.6.199>
- Brown, S. W. (1997). Attentional resources in timing: interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, 59(7), 1118–1140. <http://doi.org/10.3758/BF03205526>
- Brown, S. W. (2008). Time and attention: Review of the literature. In S. Grondin (Ed.), *Psychology of time* (pp. 111–138). Emerald.
- Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: effects of mental workload and event structure. *Journal of Experimental Psychology. Human Perception and Performance*, 28(3), 600–615.
<http://doi.org/10.1037/0096-1523.28.3.600>
- Brown, S. W., & Merchant, S. M. (2007). Processing resources in timing and sequencing tasks. *Perception & Psychophysics*, 69(3), 439–449.
<http://doi.org/10.3758/BF03193764>
- Buckner, R. L., & Wheeler, M. E. (2001). The cognitive neuroscience of remembering. *Nature Reviews. Neuroscience*, 2(9), 624–634.
<http://doi.org/10.1080/17588928.2010.503602>
- Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, 35(4), 625–641.
[http://doi.org/10.1016/S0896-6273\(02\)00830-9](http://doi.org/10.1016/S0896-6273(02)00830-9)
- Burt, C. D. B. (1999). Categorisation of action speed and estimated event duration. *Memory*, 7(3), 345–355. <http://doi.org/10.1080/096582199387968>
- Burt, C. D. B. (2002). A commentary on Pedersen and Wright. *Applied Cognitive Psychology*, 16(7), 785–787. <http://doi.org/10.1002/acp.853>
- Burt, C. D. B., & Popple, J. S. (1996). Effects of Implied Action Speed on Estimation of Event Duration. *Applied Cognitive Psychology*, 10, 53–63.
- Carlson, V. R., & Feinberg, I. (1968). Individual variations in time judgment and the concept of an internal clock. *Journal of Experimental Psychology*, 77(4), 631–40.

- Carlson, V. R., & Feinberg, I. (1970). Time judgment as a function of method, practice, and sex. *Journal of Experimental Psychology*, 85(2), 171–80.
- Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in language comprehension. *Cognitive Psychology*, 62(1), 41–79.
<http://doi.org/10.1016/j.cogpsych.2010.09.002>
- Creelman, C. (1962). Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, 34(5), 582–593.
- Dalí, S. (1931). *The Persistence of Memory* [Painting]. New York, NY: Museum of Modern Art.
- Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4), 2157–2162. <http://doi.org/10.1073/pnas.0337195100>
- Eagleman, D. M., & Pariyadath, V. (2009). Is subjective duration a signature of coding efficiency? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1525), 1841–1851.
<http://doi.org/10.1098/rstb.2009.0026>
- Efron, R. (1970). Effect of stimulus duration on perceptual onset and offset latencies. *Perception & Psychophysics*, 8(4), 231–234.
- Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, 44(1), 109–120.
<http://doi.org/10.1016/j.neuron.2004.08.028>
- Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience*, 15, 732–744.
<http://doi.org/10.1038/nrn3827>
- Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, 30, 123–152.
<http://doi.org/10.1146/annurev.neuro.30.051606.094328>
- Einstein, A. (1920). *Relativity: The special and general theory*. Penguin.
- Eisler, H. (1976). Experiments on subjective duration 1968–1975: a collection of power function exponents. *Psychological Bulletin*, 83(6), 1154–1171.
<http://doi.org/10.1037/0033-2909.83.6.1154>
- Eisler, H., Eisler, A. D., & Hellström, Å. (2008). Psychophysical issues in the study of time perception. In S. Grondin (Ed.), *Psychology of time* (pp. 75–110). Emerald.

- Ekstrom, A., & Bookheimer, S. (2007). Spatial and temporal episodic memory retrieval recruit dissociable functional networks in the human brain. *Learning & Memory*, 14, 645–654. <http://doi.org/10.1101/lm.575107>. Several
- Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. a. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. *Nature Neuroscience*, 3(11), 1149–1152. <http://doi.org/10.1038/80671>
- English, H. B., & English, A. C. (1958). *A comprehensive dictionary of psychological and psychoanalytical terms: A guide to usage*.
- Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, 81(5), 1179–1189. <http://doi.org/10.1016/j.neuron.2014.01.042>
- Fechner, G. (1860). *Elemente der Psychophysik. Elemente der psychophysik*.
- 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. <http://doi.org/10.1006/jmla.2000.2728>
- Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology. General*, 134(4), 521–537. <http://doi.org/10.1037/0096-3445.134.4.521>
- Flecken, M., & Van Bergen, G. (2014). Putting things in new places: Verb-based prediction in L1 and L2 sentence processing. In *20th Architectures and Mechanisms for Language Processing Conference (AMLAP 2014)*.
- Fraisse, P. (1948). Étude comparée de la perception et de l'estimation de la durée chez les enfants et les adultes. *Enfance*, 1(3), 199–211. <http://doi.org/10.3406/enfan.1948.1083>
- Fraisse, P. (1963). *The psychology of time*. Oxford: Harper & Row.
- Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, 35, 1-36.
- Gibbon, J., & Church, R. (1984). 26 sources of variance in an information processing theory of timing. In H.L. Roitblat et al. (Eds.) *Animal Cognition: Proceedings of the Harry Frank Guggenheim Conference*.

- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52–77. <http://doi.org/10.1111/j.1749-6632.1984.tb23417.x>
- 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. [http://doi.org/10.1016/S0959-4388\(97\)80005-0](http://doi.org/10.1016/S0959-4388(97)80005-0)
- Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition*, 52(2), 125–157. [http://doi.org/10.1016/0010-0277\(94\)90065-5](http://doi.org/10.1016/0010-0277(94)90065-5)
- Gómez, R., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4(5), 178–186. [http://doi.org/10.1016/S1364-6613\(00\)01467-4](http://doi.org/10.1016/S1364-6613(00)01467-4)
- Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin*, 127(1), 22–44.
- Grondin, S. (2008). Methods for Studying Psychological Time. In S. Grondin (Ed.), *Psychology of Time* (pp. 51–74). Emerald.
- Guyau, J. (1890). *La genèse de l'idée de temps*. Alcan.
- Hansen, J., & Trope, Y. (2013). When time flies: How abstract and concrete mental construal affect the perception of time. *Journal of Experimental Psychology: General*, 142(2), 336–347. <http://doi.org/10.1037/a0029283>
- Hanson, C., & Hirst, W. (1989). On the representation of events: a study of orientation, recall, and recognition. *Journal of Experimental Psychology: General*, 118(2), 136–147. <http://doi.org/10.1037/0096-3445.118.2.136>
- Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, 111(2), 151–167. <http://doi.org/10.1016/j.cognition.2009.01.009>
- Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 27(52), 14365–14374. <http://doi.org/10.1523/JNEUROSCI.4549-07.2007>
- Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), 1726–1731. <http://doi.org/10.1073/pnas.0610561104>

- Hellström, Å. (1977). Time errors are perceptual. *Psychological Research*, 39(4), 345-388.
- 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*, 89(4), 719-730.
- Hollingworth, H. (1909). *The inaccuracy of movement, with special reference to constant errors*. Columbia University.
- Hollingworth, H. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, 7(17), 461-469.
- Hsieh, L. T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron*, 81(5), 1165-1178. <http://doi.org/10.1016/j.neuron.2014.01.015>
- Huettel, S. A., Song, A. W., & McCarthy, G. (2009). *Functional Magnetic Resonance Imaging* (2nd ed.). Sunderland, MA: Sinauer.
- James, W. (1890). *The principles of psychology, Vol I*. New York: Holt. <http://doi.org/10.1037/10538-000>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825-841. [http://doi.org/10.1016/S1053-8119\(02\)91132-8](http://doi.org/10.1016/S1053-8119(02)91132-8)
- Joergensen, G. (2008). *Visual reflections of language-mediated internal representations of time*. University of York (UK).
- Joergensen, G., & Gennari, S. P. (2013). Exploring representations of event duration in language. Presented at 26th Annual CUNY conference on Human Sentence Processing.
- Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z., & Buckner, R. L. (2008). Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, 100(1), 129-139. <http://doi.org/10.1152/jn.00077.2008>
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kant, I. (1770). On the Form and Principles of the Sensible and Intelligible World. *Translation David Walford and Ralph Meerbote (1992)*. London: Cambridge University Press.

- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4. <http://doi.org/10.3389/neuro.06.004.2008>
- Kumaran, D., & Maguire, E. A. (2006). An unexpected sequence of events: Mismatch detection in the human hippocampus. *PLoS Biology*, 4(12), 2372–2382. <http://doi.org/10.1371/journal.pbio.0040424>
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. <http://doi.org/10.1016/j.tics.2007.11.004>
- Lehn, H., Steffenach, H.-A., van Strien, N. M., Veltman, D. J., Witter, M. P., & Håberg, A. K. (2009). A specific role of the human hippocampus in recall of temporal sequences. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 29(11), 3475–3484. <http://doi.org/10.1523/JNEUROSCI.5370-08.2009>
- Leuba, J. H. (1893). A new instrument for Weber's Law; with indications of a law of sense memory. *The American Journal of Psychology*, 5(3), 370–384. <http://doi.org/http://dx.doi.org/10.2307/1411000>
- Liverence, B. M., & Scholl, B. J. (2012). Discrete events as units of perceived time. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 549–554. <http://doi.org/10.1037/a0027228>
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585–589. [http://doi.org/10.1016/S0022-5371\(74\)80011-3](http://doi.org/10.1016/S0022-5371(74)80011-3)
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720–725. [http://doi.org/10.1016/S0193-953X\(05\)70059-9](http://doi.org/10.1016/S0193-953X(05)70059-9)
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490. <http://doi.org/10.3758/BF03210951>
- MacDonald, C. J. (2014). Prospective and retrospective duration memory in the hippocampus: is time in the foreground or background? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1637), 20120463. <http://doi.org/10.1098/rstb.2012.0463>
- MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal “time cells” bridge the gap in memory for discontinuous events. *Neuron*, 71(4), 737–749. <http://doi.org/10.1016/j.neuron.2011.07.012>

- Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35(8), 1489–1517. <http://doi.org/10.1111/j.1551-6709.2011.01202.x>
- Mankin, E. A, Sparks, F. T., Slayyeh, B., Sutherland, R. J., Leutgeb, S., & Leutgeb, J. K. (2012). Neuronal code for extended time in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47), 19462–7. <http://doi.org/10.1073/pnas.1214107109>
- Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual Changes in Hippocampal Activity Support Remembering the Order of Events. *Neuron*, 56(3), 530–540. <http://doi.org/10.1016/j.neuron.2007.08.017>
- Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: Coincidence detection of oscillatory processes. *Cognitive Brain Research*, 21(2), 139–170. <http://doi.org/10.1016/j.cogbrainres.2004.06.012>
- Matthews, W., Stewart, N., & Wearden, J. (2011). Stimulus Intensity and the Perception of Duration. *Journal of Experimental Psychology: Human Perception and Performance*. 31(1), 303–313.
- Mayes, A R., Isaac, C. L., Holdstock, J. S., Hunkin, N. M., Montaldi, D., Downes, J. J., ... Roberts, J. N. (2001). Memory for single items, word pairs, and temporal order of different kinds in a patient with selective hippocampal lesions. *Cognitive Neuropsychology*, 18(2), 97–123. <http://doi.org/10.1080/02643290125897>
- McClain, L. (1983). Interval estimation: effect of processing demands on prospective and retrospective reports. *Perception & Psychophysics*, 34(2), 185–189. <http://doi.org/10.3758/BF03211347>
- 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.
- Murphy, G. (2002). *The big book of concepts*. MIT Press.
- Naya, Y., & Suzuki, W. A. (2011). Integrating what and when across the primate medial temporal lobe. *Science*, 333(6043), 773–776. <http://doi.org/10.1126/science.1206773>
- Newton, D. (1976). Foundations of attribution: The perception of ongoing behavior. In J. H. Harvey, W. J. Ickes, & R. Kidd (Eds.), *New directions in attribution research*. (pp. 223–248). Hillsdale: Erlbaum.
- Newton, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12(5), 436–450. [http://doi.org/10.1016/0022-1031\(76\)90076-7](http://doi.org/10.1016/0022-1031(76)90076-7)

- Nichols, H. (1891). The Psychology of Time. *The American Journal of Psychology*, 3(4), 453–529.
- Nielson, J. (1958). *Memory and amnesia*. San Lucas Press.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon.
- Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., ... Yamadori, A. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *NeuroImage*, 19(4), 1369–1380. [http://doi.org/10.1016/S1053-8119\(03\)00179-4](http://doi.org/10.1016/S1053-8119(03)00179-4)
- Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), 2745–2750. <http://doi.org/10.1073/pnas.0708424105>
- Ornstein, R. E. (1969). *On the experience of time*. Penguin Books.
- Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS ONE*, 2(11), 2–7. <http://doi.org/10.1371/journal.pone.0001264>
- Paz, R., Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2010). A neural substrate in the human hippocampus for linking successive events. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13), 6046–6051. <http://doi.org/10.1073/pnas.0910834107>
- Pedersen, A. C. I., & Wright, D. B. (2002). Do differences in event descriptions cause differences in duration estimates? *Applied Cognitive Psychology*, 16(7), 769–783. <http://doi.org/10.1002/acp.827>
- Poynter, W. D. (1979). *Human time perception and memory processes: the role of retrieval in duration estimation*. Arizona State University.
- Poynter, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition*, 11(1), 77–82. <http://doi.org/10.3758/BF03197664>
- Poynter, W. D. (1989). Judging the duration of time intervals: a process of remembering segments of experience. *Advances in Psychology*, (59), 305–331. [http://doi.org/10.1016/S0166-4115\(08\)61045-6](http://doi.org/10.1016/S0166-4115(08)61045-6)
- Poynter, W. D., & Homa, D. (1983). Duration judgement and the experience of change. *Perception & Psychophysics*, 33(6), 548–560.

- 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. [http://doi.org/10.1016/0001-6918\(95\)00044-5](http://doi.org/10.1016/0001-6918(95)00044-5)
- Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 608–620. <http://doi.org/10.1002/wcs.133>
- Roeckelein, J. E. (2008). History of conceptions and accounts of time and early time perception research. In S. Grondin (Ed.), *Psychology of time* (pp. 1–50). Emerald.
- Rosenbaum, R. S., Priselac, S., Köhler, S., Black, S. E., Gao, F., Nadel, L., & Moscovitch, M. (2000). Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. *Nature Neuroscience*, 3(10), 1044–1048. <http://doi.org/10.1038/79867>
- Sahakyan, L., & Smith, J. R. (2014). “A long time ago, in a context far, far away”: retrospective time estimates and internal context change. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 86–93. <http://doi.org/10.1037/a0034250>
- Sawyer, T. F., Meyers, P. J., & Huser, S. J. (1994). Contrasting task demands alter the perceived duration of brief time intervals. *Perception & Psychophysics*, 56(6), 649–657. <http://doi.org/10.3758/BF03208358>
- Schacter, D. L., & Addis, D. R. (2009). On the nature of medial temporal lobe contributions to the constructive simulation of future events. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1521), 1245–1253. <http://doi.org/10.1098/rstb.2008.0308>
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews. Neuroscience*, 8(9), 657–661. <http://doi.org/10.1080/08995600802554748>
- Schapiro, A. C., Gregory, E., & Landau, B. (2014). The Necessity of the Medial-Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience*, 26(8), 1736–1747. <http://doi.org/10.1162/jocn>
- Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, 22(17), 1622–1627. <http://doi.org/10.1016/j.cub.2012.06.056>

- Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, 16(4), 486–92. <http://doi.org/10.1038/nn.3331>
- Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, 37(6), 1013–1025. [http://doi.org/10.1016/S0896-6273\(03\)00123-5](http://doi.org/10.1016/S0896-6273(03)00123-5)
- Schiffman, H. R., & Bobko, D. J. (1974). Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology*, 103(1), 156–159. <http://doi.org/10.1037/h0036794>
- Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology*, 18(1), 37–55. <http://doi.org/10.1002/acp.940>
- Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition*, 28(2), 214–223. <http://doi.org/10.3758/BF03213801>
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20(1), 11–21.
- Shastri, L. (2002). Episodic memory and cortico – hippocampal interactions. *Trends in Cognitive Sciences*, 6(4), 162–168.
- Sirigu, A., & Grafman, J. (1996). Selective impairments within episodic memories. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 32(1), 83–95. [http://doi.org/10.1016/S0010-9452\(96\)80018-9](http://doi.org/10.1016/S0010-9452(96)80018-9)
- Slotnick, S. D., & Schacter, D. L. (2006). The nature of memory related activity in early visual areas. *Neuropsychologia*, 44(14), 2874–2886. <http://doi.org/10.1016/j.neuropsychologia.2006.06.021>
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6), 246–251. [http://doi.org/10.1016/S1364-6613\(03\)00109-8](http://doi.org/10.1016/S1364-6613(03)00109-8)
- Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activations neural representations of visual and motor experiences. *Psychological Science*, 20(8), 289–299.

- Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2004). Perceiving narrated events. In K. Forbus et al. (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (p. 1637). Chicago.
- Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, 18(5), 449–55.
<http://doi.org/10.1111/j.1467-9280.2007.01920.x>
- Spiers, H. J., & Maguire, E. A. (2007). Decoding human brain activity during real-world experiences. *Trends in Cognitive Sciences*, 11(8), 356–365.
<http://doi.org/10.1016/j.tics.2007.06.002>
- Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science (New York, N.Y.)*, 253(5026), 1380–1386.
<http://doi.org/10.1126/science.1896849>
- Staresina, B. P., & Davachi, L. (2009). Mind the Gap: Binding Experiences across Space and Time in the Human Hippocampus. *Neuron*, 63(2), 267–276.
<http://doi.org/10.1016/j.neuron.2009.06.024>
- Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, 23(5), 1052–1064.
<http://doi.org/10.1162/jocn.2010.21524>
- 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. <http://doi.org/10.1371/journal.pone.0009271>
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock”. *Psychological Monographs: General and Applied*. 7(13), 1-31.
- Tubridy, S., & Davachi, L. (2011). Medial temporal lobe contributions to episodic sequence encoding. *Cerebral Cortex*, 21(2), 272–280.
<http://doi.org/10.1093/cercor/bhq092>
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). New York: Academic. Retrieved from <http://doi.apa.org/psycinfo/1973-08477-007>
- Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Clarendon.
- Tulving, E. (1984). Relations among components and processes of memory. *Behavioral and Brain Sciences*, 7(2), 257–268.
<http://doi.org/10.1017/S0140525X00044617>

- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, 26(1), 1–12.
- Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, 53(1), 1–25.
<http://doi.org/10.1146/annurev.psych.53.100901.135114>
- Tulving, E., Schacter, D. L., McLachlan, D. R., & Moscovitch, M. (1988). Priming of semantic autobiographical knowledge: a case study of retrograde amnesia. *Brain and cognition*, 8(1), 3–20. [http://doi.org/10.1016/0278-2626\(88\)90035-8](http://doi.org/10.1016/0278-2626(88)90035-8)
- 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.
<http://doi.org/10.1126/science.277.5324.376>
- Vroon, P. A. (1970). Effects of presented and processed information on duration experience. *Acta Psychologica*, 34, 115–121. [http://doi.org/10.1016/0001-6918\(70\)90010-7](http://doi.org/10.1016/0001-6918(70)90010-7)
- Wagner, A. D., Wagner, A. D., Shannon, B. J., Shannon, B. J., Kahn, I., Kahn, I., ... Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, 9(9), 445–53.
<http://doi.org/10.1016/j.tics.2005.07.001>
- Waldum, E. R., & Sahakyan, L. (2013). A Role for Memory in Prospective Timing Informs Timing in Prospective Memory. *Journal of Experimental Psychology: General*, 142(3), 809–826. <http://doi.org/10.1037/a0030113>
- Wearden, J. H. (1991). Do humans possess an internal clock with scalar timing properties? *Learning and Motivation*, 22(1-2), 59–83.
[http://doi.org/10.1016/0023-9690\(91\)90017-3](http://doi.org/10.1016/0023-9690(91)90017-3)
- 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.
- Weber, E. (1834). *De Pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae*.
- Wiener, M., Turkeltaub, P., & Coslett, H. B. (2010). The image of time: A voxel-wise meta-analysis. *NeuroImage*, 49(2), 1728–1740.
<http://doi.org/10.1016/j.neuroimage.2009.09.064>

- Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Van den Bosch, A. (in press). Prediction during natural language comprehension. *Cerebral Cortex*, (in press).
- 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 of the United States of America*, 104(19), 7780–7785. <http://doi.org/10.1073/pnas.0701644104>
- Woehrle, J. L., & Magliano, J. P. (2012). Time flies faster if a person has a high working-memory capacity. *Acta Psychologica*, 139(2), 314–319. <http://doi.org/10.1016/j.actpsy.2011.12.006>
- Woodrow, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, 13, 473–499.
- Woodrow, H. (1933). Individual differences in the reproduction of temporal intervals. *American Journal of Psychology*, 45, 271–281.
- Woodrow, H. (1934). The temporal indifference interval determined by the method of mean error. *Journal of Experimental Psychology*, 17, 167–188.
- Woodrow, H. (1951). Time perception. In *Handbook of experimental psychology* (pp. 1224–1236).
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, 14(6), 1370–1386. <http://doi.org/10.1006/nimg.2001.0931>
- Wundt, W. (2009). Outlines of psychology (1897). In *Foundations of psychological thought: A history of psychology* (pp. 36–44).
- Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 356(1413), 1363–74.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441–517. <http://doi.org/10.1006/jmla.2002.2864>
- Yonelinas, A. P., & Jacoby, L. L. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language*, 34(5), 622–643. <http://doi.org/10.1006/jmla.1995.1028>

- Yonelinas, A. P., Kroll, N. E. A., Quamme, J. R., Lazzara, M. M., Sauvé, M.-J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, 5(11), 1236–1241. <http://doi.org/10.1038/nn961>
- Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 25(11), 3002–3008. <http://doi.org/10.1523/JNEUROSCI.5295-04.2005>
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979–1008. <http://doi.org/10.1016/j.cogsci.2004.06.003>
- 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. <http://doi.org/10.1038/88486>
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology. General*, 138(2), 307–327. <http://doi.org/10.1037/a0015305>
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, 133(2), 273–293. <http://doi.org/10.1037/0033-2909.133.2.273>
- 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), 466–482. <http://doi.org/10.1037/0882-7974.21.3.466>
- Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16(2), 80–84. <http://doi.org/10.1111/j.1467-8721.2007.00480.x>
- Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, 1076(1), 150–162. <http://doi.org/10.1016/j.brainres.2005.12.122>
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127(1), 3–21.
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology. General*, 130(1), 29–58. <http://doi.org/10.1037/0096-3445.130.1.29>

- Zakay, D. (1992). On prospective time estimation, temporal relevance and temporal uncertainty. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap*. Dordrecht, Netherlands: Kluwer Academic.
- Zakay, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics*, 54(5), 656–664. <http://doi.org/10.3758/BF03211789>
- Zakay, D., & Block, R. A. (1997). Temporal Cognition. *Current Directions in Psychological Science*, 6(1), 12–16.
- Zakay, D., & Block, R. A. (2004). Prospective and retrospective duration judgments: An executive-control perspective. *Acta Neurobiologiae Experimentalis*, 64(3), 319–328.
- 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. <http://doi.org/10.3758/BF03200861>
- Zelkind, I. (1973). Factors in time estimation and a case for the internal clock. *The Journal of General Psychology*, 88(2), 295–301.
- Zimbardo, P. G., & Boyd, J. N. (1999). Putting time in perspective: A valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, 7(6), 1291–1288. <http://doi.org/10.1037//0022-3514.77.6.1271>
- 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), 386–397.